ISSN 2328-224X

Journal of Statistical Science and Application

Volume 2, Number 3, March 2014

DAVID PUBLISHING

From Knowledge to Wisdom

David Publishing Company
www.davidpublishing.com

Journal of Statistical Science and Application

Volume 2, Number 3, March 2014 (Serial Number 4)

David Publishing Company
www.davidpublishing.org

Publication Information:
Journal of Statistical Science and Application is published monthly in hard copy and online (ISSN 2328-224X) by David Publishing Company located at 240 Nagle Avenue #15C, New York, NY 10034, USA.

Aims and Scope:
Journal of Statistical Science and Application, a monthly professional academic and peer-reviewed journal, particularly emphasizes new research results in theory and methodology, in realm of Mathematical Statistics, Applied Statistics, Decision Analysis, Probability Theory & Applications, Statistical Computing, Biostatistics, Business Statistics, Chemometrics, Data Mining, Econometrics, Engineering Statistics, Psychological Statistics, Sports Statistics, etc.. Articles interpreting practical application of up-to-date technology and method are also welcome.

Manuscripts and correspondence are invited for publication. You can submit your papers via web submission, or E-mail to statistical@davidpublishing.org, jssa.statistical@hotmail.com, jssa.statistical@gmail.com. Submission guidelines and web submission system are available at http://www.davidpublishing.org or http://www.davidpublishing.com.

Editorial Office:
240 Nagle Avenue #15C, New York, NY 10034, USA.
Tel: 1-323-984-7526, 323-410-1082
Fax: 1-323-984-7374, 323-908-0457
E-mail: statistical@davidpublishing.org, jssa.statistical@hotmail.com, jssa.statistical@gmail.com

Subscription Information:
Price (per year):
Print $520; Online $320; Print and Online $560

David Publishing Company
240 Nagle Avenue #15C, New York, NY 10034, USA.
Tel: 1-323-984-7526, 323-410-1082; Fax: 1-323-984-7374, 323-908-0457
E-mail: order@davidpublishing.com

DAVID PUBLISHING

David Publishing Company
www.davidpublishing.com

Journal of Statistical Science and Application

Volume 2, Number 3, March 2014 (Serial Number 4)

Contents

Journal of Statistical Science and Application 2 (2014) 85-92

Practical Aspects for Designing Statistically Optimal Experiments

Mark J. Anderson, Patrick J. Whitcomb
Stat-Ease, Inc., Minneapolis, MN USA

Due to operational or physical considerations, standard factorial and response surface method (RSM) design of experiments (DOE) often prove to be unsuitable. In such cases a computer-generated statistically-optimal design fills the breech. This article explores vital mathematical properties for evaluating alternative designs with a focus on what is really important for industrial experimenters. To assess "goodness of design" such evaluations must consider the model choice, specific optimality criteria (in particular D and I), precision of estimation based on the fraction of design space (FDS), the number of runs to achieve required precision, lack-of-fit testing, and so forth. With a focus on RSM, all these issues are considered at a practical level, keeping engineers and scientists in mind. This brings to the forefront such considerations as subject-matter knowledge from first principles and experience, factor choice and the feasibility of the experiment design.

Key words: design of experiments, optimal design, response surface methods, fraction of design space.

Introduction

Statistically optimal designs emerged over a half century ago (Kiefer, 1959) to provide these advantages over classical templates for factorials and RSM:

- Efficiently filling out an irregularly shaped experimental region such as that shown in Figure 1 (Anderson and Whitcomb, 2005),

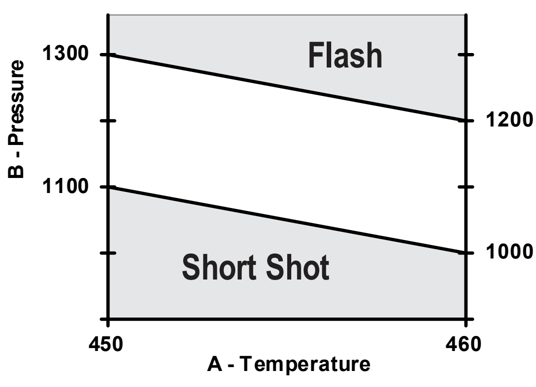

Figure 1. Example of irregularly-shaped experimental region (a molding process).

- Minimizing the runs to just what is needed to fit the assumed polynomial model,
- Accommodating unusual requirements concerning either the number of blocks or the number of runs per

Corresponding author: Mark J. Anderson, Stat-Ease, Inc., Minneapolis, MN USA. E-mail: mark@statease.com.

block,

- Handling a combination of factor types, such as continuous, discrete, categorical and mixture.

Over time a number of criteria labeled alphabetically became favored by industrial experiments for optimal designs (Box and Draper, 2007), of which two will be covered in this article: I-optimal and D-optimal. Our focus will be kept to RSM.

What is a "Good" Experiment Design?

To answer this important question, let's start with a wish-list for choosing a "suitable" experiment design derived from one provided by George Box (Box, 1982)—the co-inventor of RSM (Box and Wilson, 1951):

(1) Allow the chosen polynomial to be estimated well.

(2) Give sufficient information to allow a test for lack of fit by

a. Having more unique design points than model coefficients, and

b. Providing an estimate of "pure" error, i.e., replicates.

(3) Remain insensitive to outliers, influential values and bias from model misspecification.

(4) Be robust to errors in control of the factor levels.

(5) Permit blocking and sequential experimentation.

(6) Provide a check on variance assumptions, e.g., studentized residuals are normal with a mean of zero and constant variance.

(7) Generate useful information throughout the region of interest, i.e., provide a good distribution of standard error of predictions.

(8) Do not contain an excessively large number of trials.

When applying RSM, industrial experimenters generally choose as step one a quadratic polynomial, which are remarkably versatile for empirical modeling. For this purpose, the central composite design (CCD), also known as the Box-Wilson in honor of the developers, scores well on all the desired attributes. However, standard layouts like the CCD are not a good fit for non-cuboidal regions such as that illustrated in Figure 1. For situations like this or others spelled out in the Introduction, optimal designs are seemingly the panacea. However, as we will discuss further, you had best keep in mind that *"designing an experiment should involve balancing multiple objectives, not just focusing on a single characteristic"* (Myers, Montgomery and Anderson-Cook, 2009).

Purely Optimal Designs—Comparing I Versus D as the Criterion

Although there are many variations on the theme of optimal design, the following two criteria are the ones primarily used in industrial experimentation:

- optimal (also known as "IV") to minimize the integral of the prediction variance
- D-optimal to minimize the volume of the confidence ellipsoid for the coefficients and thus maximize information on the polynomial coefficients.

Rather than getting mired down in the mathematical details for these two criteria and the multitude of algorithms for applying them (readily available from a vast array of references—some already cited here), we will focus on how they differ in actual application to RSM experiments designed to fit quadratic polynomials such as the equation show above. For this purpose a good tool for comparison is the standard error (SE) plot, such as those shown in Figure 2 for a 12-run RSM design on one factor where the points are picked I- versus D-optimally.

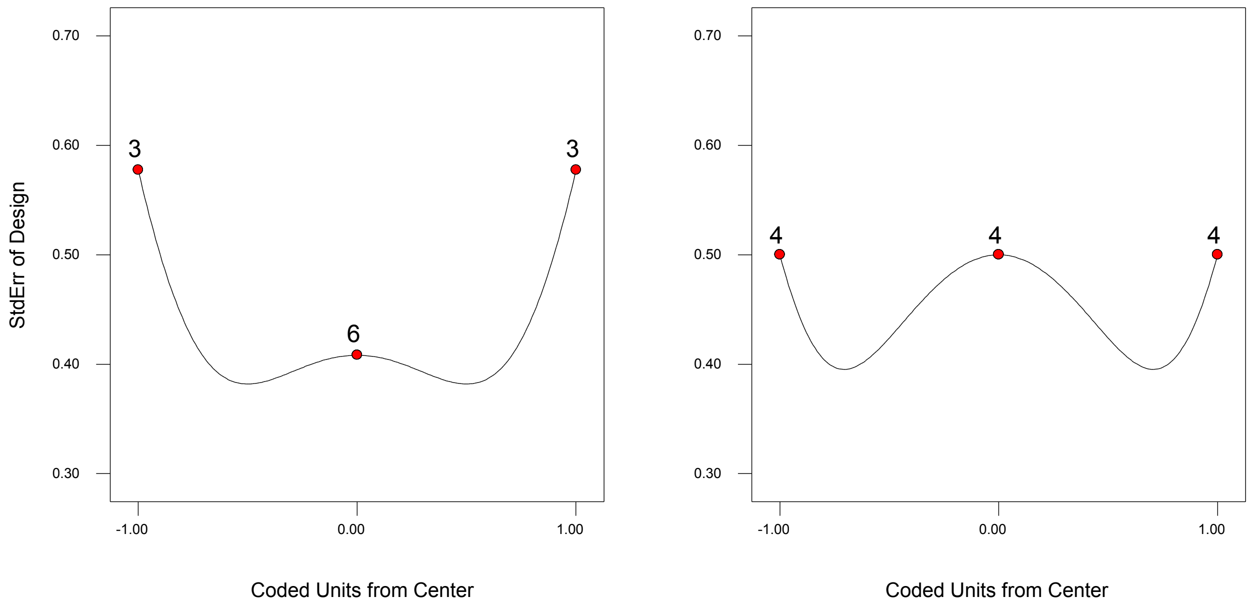

Figure 2. Standard error plots for I (left) versus D optimal (right) for a 12-run, one-factor RSM design.

Comparing these plots side-by-side provides a picture of the relative quality of predicted response at various locations spanning the experimental region (shown here in terms of coded units from the center). The greater replication of points by the I-criterion is desirable for RSM because it lowers the standard error of prediction at the center—the point of greatest interest—and provides a fairly flat profile for a broader (relative to the D-criterion) range in the middle of the experimental region.

Another way to compare designs is via the fraction of design space (FDS) plot (Anderson-Cook, Borror and Montgomery, 2009), which consists of a single line for a given design, thus allowing display of the prediction variance for several designs at once. Figure 3 lays out the FDS curves for the aforementioned 12-run RSM design done by the competing criterion—I versus D.

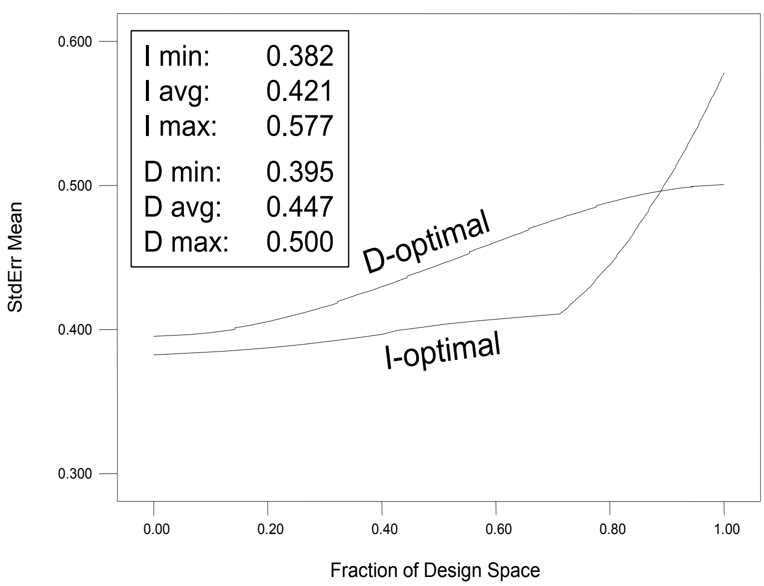

Figure 3. FDS plots for I versus D optimal for a 12-run, one-factor RSM design.

The legend in Figure 3 provides benchmarks on the standard error minimums, averages and maximums. For RSM purposes, the I-optimal provides a desirable tradeoff of being higher at its maximum (not good) but lower on average (good) than the D-optimal design.

At this stage we put I-optimal at the forefront. However, neither of the two optimal designs under consideration provide for a lack-of-fit (LOF) test. Therefore an experimenter cannot assess whether the model they chose provides an adequate approximation of the true response. It's vital to keep in mind that *"no postulated model can ever be assumed to be fully correct [therefore] the basic assumptions underlying the alphabetic-optimality approach are often unrealistic <u>from the practical viewpoint</u> of actually designing real experiments"* (Draper & Guttman, 1988).

Modifying Optimal Designs to Make Them More Robust to Model Misspecification

We will now do another comparison of point selection using I- vs D-optimal on a one-factor design for a quadratic model, but this time the designs will be modified with LOF points to check for model misspecification. We recommend 4 runs for these 'check-points'—chosen to maximize the minimum distance from existing design points; thus filling 'holes' in the experimental space. This is known as the "distance" criterion. Figure 4 provides a side-by-side view, by way of standard error plots, of how this modification affects the spread of points compared to the purely optimal selection shown in Figure 2.

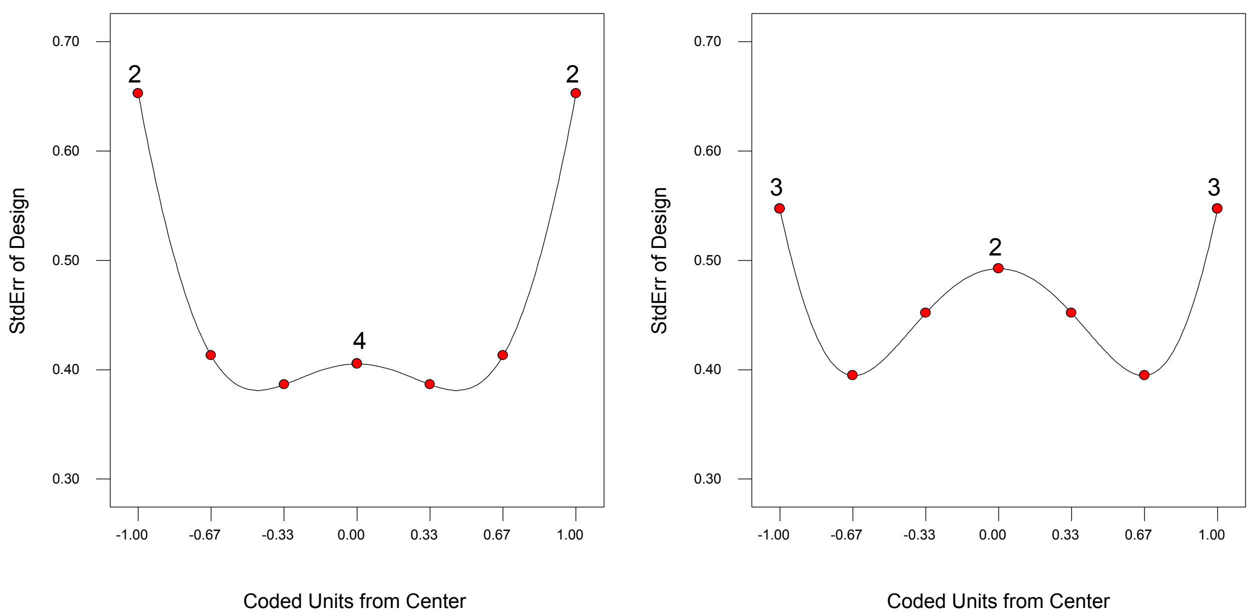

Figure 4. SE plots for LOF-modified I-optimal (left) versus LOF-modified D-optimal (right) for a 12-run, one-factor RSM design.

Observe how gaps in the experimental range have been plugged.

Next let's recalculate the FDS for the two alternative optimal criterion.

Compare and contrast the curves and data on Figures 3 and 5—the first being purely optimal and the second modified with lack-of-fit points. Notice that by all statistical measures and the curves themselves that not much differs. In this case the advantage of having a check for model misspecification outweighs the minor loss in optimality and slight degradation in FDS quality.

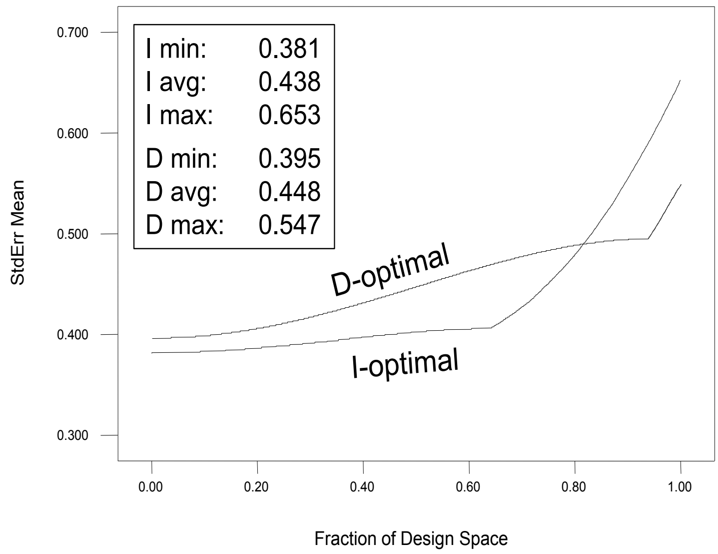

Figure 5. FDS plots for LOF-modified I versus D optimal for a 12-run, one-factor RSM design

Extending These Findings to Two Factors

A similar case can be made for two factors and, by extension, beyond—once the minimum points needed to fit the chosen polynomial model are selected via an optimal criterion, adding LOF points causes little harm and they create a lot of good. Figure 6 lays out the points on SE plot for a 14-run, I-optimal, RSM-quadratic design with zero versus four LOF points.

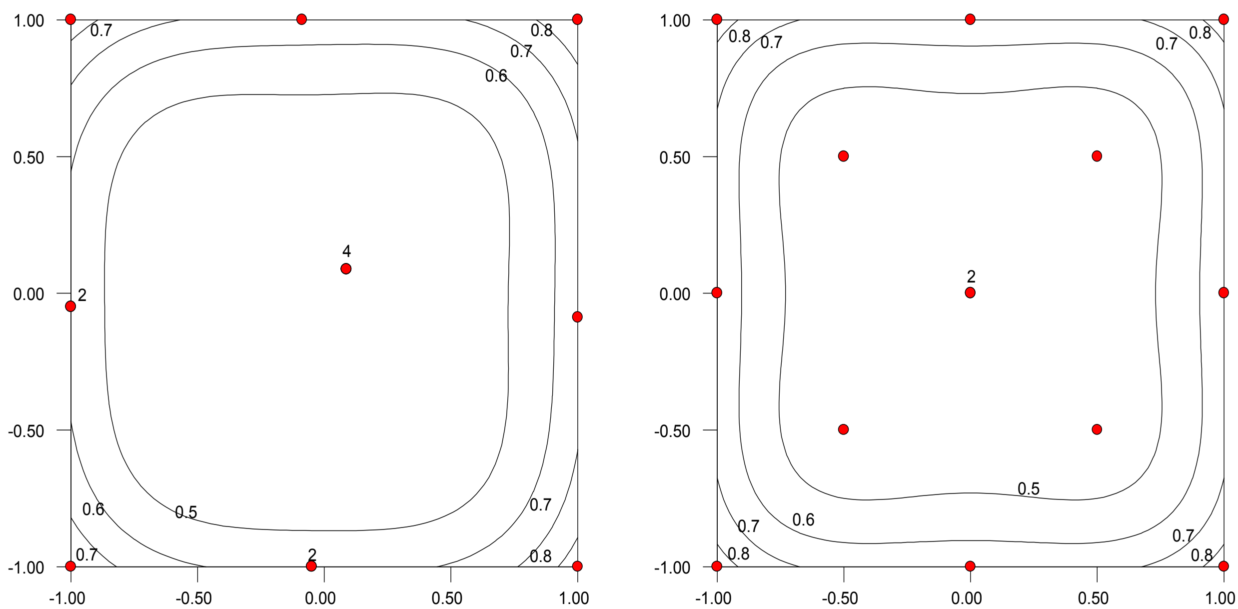

Figure 6. SE plots for purely I-optimal (left) versus LOF-modified (right) for a 14-run, two-factor RSM design.

The space-filling effect of shifting some points from optimal to distance-based criterion (for LOF) is very apparent when comparing these plots from left to right. This comes at little cost to the I-optimal design quality as evidenced by the FDS curve and properties laid out in Figure 7.

However, when we make the same comparison for D-optimal designs, the trade-off of 4 purely optimal points for ones chosen by distance (layouts shown in Figure 8) to provide LOF does not go quite as well—see how the FDS curve shifts in Figure 9 and note the degradation in D-optimality as evidenced by the determinant results being increased.

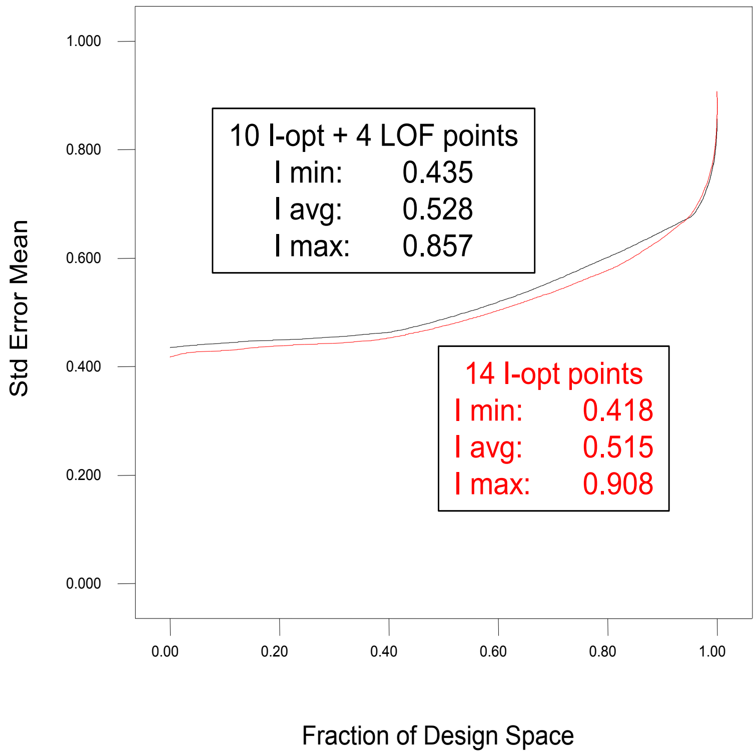

Figure 7. FDS plots for purely I-optimal versus LOF-modified I-optimal for a 14-run, two-factor RSM design.

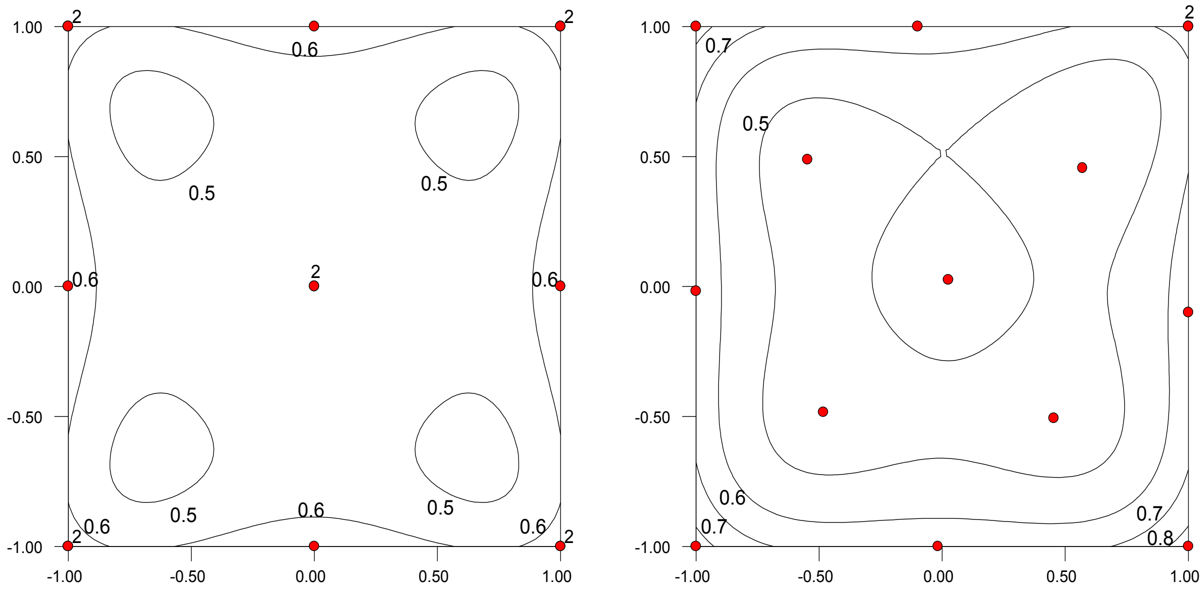

Figure 8. SE plots for purely D-optimal (left) versus LOF-modified D-optimal (right) for a 14-run, two-factor RSM design.

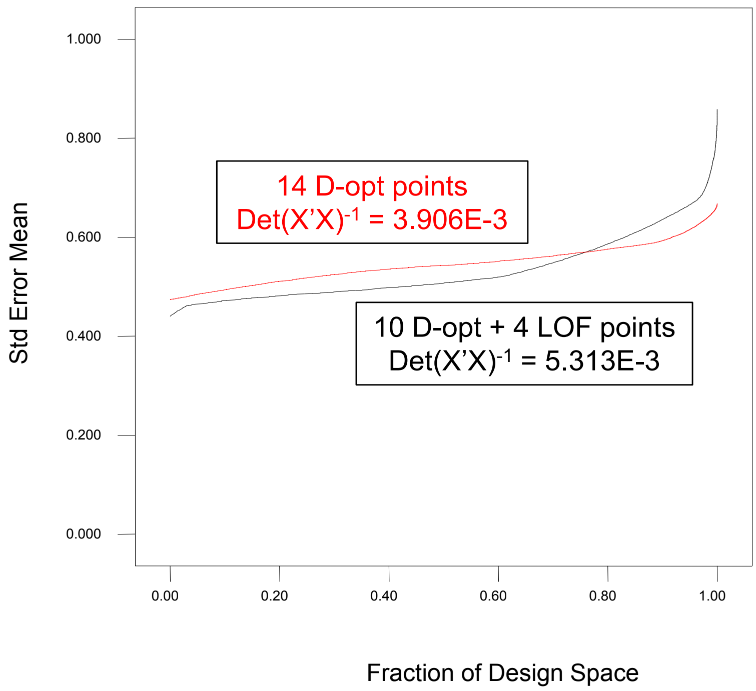

Figure 9. FDS plots for purely D-optimal versus LOF-modified D-optimal for a 14-run, two-factor RSM design.

This lends further support for the use of I over D-optimal designs for response surface method optimization.

Bolstering the Design with Replicates

Replicates are needed to provide the pure error needed for a LOF test. In the cases presented above they came with the optimal portion of the designs. A better way to build a practical experiment is to detail the number of replicates and choose them by the optimal criterion—we recommend a minimum of 4 points being selected in this manner. This provides enough degrees of freedom for a reasonably-powerful LOF test.

Conclusion

In consideration of the practical aspects of algorithmic design for RSM we recommend adding by distance-base at least 4 points for testing lack of fit, even though this makes the experiment less alphabetically optimal. This a good trade off. Furthermore, in physical experiments it is desirable build in an estimate of experimental error based on at least 4 degrees of freedom via replicated design point(s). For choice of optimal criterion we advise using I- for empirical modeling (RSM)—reserving D-optimal (because of their more-precise estimation of model coefficients) for factorial screening.

However, even with these practical aspects incorporated, optimal design cannot make up for:

- Choosing the factors that are not among the vital few.
- Designing for an inadequate empirical model.
- Going out of the region of operability.
- Measuring the wrong responses.

> "The exact functional relationship is usually unknown and possibly unknowable. We have only to think of the flight of a bird, the fall of a leaf, or the flow of water through a valve to realize that we are likely to be able to approximate only the main features of such a relationship." – Box & Draper

References

Anderson, M. J. and Whitcomb, P.J. (2007) RSM Simplified, *Optimizing Processes Using Response Surface Methods for Designs of Experiments*, Computer-Generated Optimal Designs," 7: 149-150, New York: Productivity Press.

Anderson-Cook, C.M., Borror, C. M. and Montgomery, D. C. (2009) "Response surface design evaluation and comparison," *Journal of Statistical Planning and Inference*, 139: 629–641.

Box, G. E. P., and Wilson, K B. (1951) "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society, Series B* 13:1-45.

Box, G. E. P. (1982) "Choice of Response Surface Design and Alphabetic Optimality," *Utilitas Mathematica*, 21B: 11-55.

Box, G. E. P., and Draper, N. R. (2007) *Response Surfaces, Mixtures, and Ridge Analyses*, "The Alphabetic Optimality Approach," 14.4: 471-481, New York: Wiley.

Draper, N. R. and Guttman, I. (1988) "Rationalization of the "Alphabetic-Optimal" and "Variance Plus Bias" Approaches to Experimental Design," Technical Report 841, Department of Statistics, University of Wisconsin.

Kiefer, J. (1959). "Optimum Experimental Designs," *Journal of the Royal Statistical Society, Series B* 21: 272–319.

Myers, R. H., Montgomery, D. C. and Anderson-Cook, C.M. (2009) *Response Surface Methodology, 3rd Edition*, New York: Wiley.

Journal of Statistical Science and Application 2 (2014) 93-101

Application of Statistical Methods and GIS for Downscaling and Mapping Crop Statistics Using Hypertemporal Remote Sensing

Ahmed DOUAIK
Research Unit on Environment and Conservation of Natural Resources,
National Institute of Agricultural Research (INRA), Rabat, Morocco. Email: ahmed_douaik@yahoo.com

To sustain the management of natural resources, land use and land cover (LULC) should be spatially mapped and temporally monitored using GIS. For large areas, conventional methods are laborious. Alternatively, remote sensing can be used for LULC mapping and monitoring. Normalized differential vegetation index (NDVI) is the most used vegetation index for crop identification and phenology. For agricultural areas, crop statistics are estimated yearly at regional level following administrative units. However, these statistics are not informing about spatial extent of these crops within administrative units; such information is crucial for crop monitoring. The main objective of this research was to fill the gap, based on statistical methods and GIS, by adding spatial information to crop statistics by analyzing temporal NDVI profiles. The study area covers 1300 km^2. Data consist of 147 decadal Spot Vegetation NDVI images. Crop statistics were compiled on seasonal basis and aggregated to different administrative levels. Images were processed using an unsupervised classification method. A series of classification runs corresponding to different numbers of clusters were used. Using stepwise multiple linear regression, cropped areas from agricultural statistics were related to areas of each NDVI profile cluster. Estimated regression coefficients were used to generate maps showing cropped fractions by map units. The optimal number of clusters was 18. Similar profiles were merged leading to eight clusters. The results show that, for example, rice was grown, in autumn, on 50% of the area of map-units represented by NDVI-profile group 4 and 75% of the area of group 7 while it was grown, in spring, on 2, 69 and 25% of areas of NDVI-profile groups 2, 6, and 7, respectively. Regression coefficients were used to generate map of crops. This research illustrates the benefit of integrating statistical methods, GIS, remote sensing and crop statistics to delineate NDVI profile clusters with their corresponding agricultural land cover map units and to link these statistics to geographical locations. These map units can be used as a reference for future monitoring of natural resources, in particular crop growth and development and for forecasting crop production and/or yield and stresses like drought.

Keywords: Crop Statistics; GIS; Multiple Regression; NDVI; Unsupervised Classification.

Introduction

The food needs of the ever increasing world population should be satisfied quantitatively and qualitatively. Since the spatial extent of arable lands is limited, the focus is currently on a better and sustainable use and management of natural resources, including soil and land resources. In order to attain this sustainability, land

---

Corresponding author: Ahmed DOUAIK, Research Unit on Environment and Conservation of Natural Resources, National Institute of Agricultural Research (INRA), Rabat, Morocco. Email: ahmed_douaik@yahoo.com.

use and land cover should be spatially mapped and temporally monitored. As the areas to be mapped are very large, conventional methods, through aerial photo-interpretation are laborious and expensive (Tucker, 1979; Philipson, 1997; Taylor et al, 2000; Falkner and Dennis, 2002). Alternatively, satellite remote sensing tools can be beneficially used for land use and land cover mapping and monitoring (Cihlar, 2000; Lillesand et al, 2007; Giri, 2012). In this way, the normalized differential vegetation index (NDVI), initially proposed by Rouse et al (1973) and which measures the vigor and greenness of vegetation (Tarpley et al, 1984), is the most used among the vegetation indices for studying vegetation, and specifically crop, phenologies (Sarkar and Kafatos, 2004; Sakamoto et al, 2005; White et al, 2009; Atkinson et al, 2012; You et al, 2013) and yield forecasting (Das et al, 1993; Ferencz et al, 2004; Mkhabela et al, 2005). Remote sensing was also used for estimating cropped area (Campbell et al, 1987; Sheub and Atkins, 1991; Labus et al, 2002; Howard et al, 2012). Time series of NDVI were used to discriminate between vegetation and other land uses (Nordberg and Evertson, 2003; Knight et al, 2006) and, for vegetation, between different green areas specific to crops, forests, etc (Murakami and al, 2001; Balaghi et al, 2008; Xie et al, 2008). For agricultural areas, crop statistics (mainly cropped areas and production) are estimated yearly at regional level following given administrative units (USDA, 2014). However, these statistics are not informing about the spatial extent of these crops within administrative units, such information is crucial for crop monitoring in future time. Since more than two decades ago, remote sensing was helpful in determining crop acreage and productivity (Allen, 1990; Gonzalez-Alonso et al, 1997; Carfagna and Gallego, 2005). However, it is just very recently that hypertemporal remote sensing was used to describe and map variability of cropping patterns of different crops in Spain (Khan et al, 2010), rice in Vietnam (Nguyen et al, 2012), winter crops in Australia (Potgieter et al, 2013), relationship between the fraction of evergreen forests and the presence of epiphyllous liverworts in China (Jiang et al, 2013), gradient in land cover vegetation growth in Greece (Ali et al, 2013), natural landscape heterogeneity in Greece (Ali et al, 2014). The main objective of this research work was to fill the gap in the official crop statistics by adding them the spatial information through the analysis of hypertemporal remote sensing, i.e., temporal NDVI profiles using different statistical methods.

Material and Methods

Study area

The study area is situated in the western part of Nizamabad district, Hyderabad province, Andra Pradesh state, in central India. The district has an irrigation system used for rice cultivation, cotton cultivation on vertisols, dryland cropping on poor sandy soils, forests on hilly land, and degraded areas. The study area is spatially very heterogeneous. The soils are classified into four main orders: Inceptisols, Alfisols, Vertisols, and Entisols. The climate is tropical with hot summers (maximum mean monthly temperature of about 40 °C) and cool and dry winters (maximum mean monthly temperature of about 13 °C). Regarding rainfall, it is about 900 mm which occur during 2 months during the southwest monsoon. Six Mandals or sub-districts are concerned by this study covering 1300 km^2 from which 90000 Ha are agricultural lands and 18000 Ha are shrub and non cultivated area.

Data

Data consist of 147 geo-referenced and stacked Spot Vegetation composite NDVI images provided by VITO (http://www.VGT.vito.be). They have a spatial resolution of 1 km^2 and available on a decadal basis for a

period ranging from April 1998 to April 2002. Two other types of data were used. Land cover map at 1/50000 scale was established from images acquired in 1994/1995 by the Indian remote sensing satellite IRS-C using the Liss-III sensor (spatial resolution of 23 m). For this work the original 18 legend entries were simplified and reduced to seven. Regarding crop statistics, they were compiled on seasonal basis and aggregated to different administrative levels (CPO, 2001).

Methods

The normalized differential vegetation index is defined by:

$$NDVI = (IR - R) / (IR + R) \qquad \text{Eq (1)}$$

IR and R are the infra red (0.78 – 0.89 μm) and red (0.61 – 0.68 μm) bands, respectively, which are bands 3 and 2 for Spot Vegetation.

The NDVI values were reported as digital number (DN) values, ranging between 0 and 255, using the following equation:

$$DN = (NDVI + 0.1) / 0.004 \qquad \text{Eq (2)}$$

The stacked 147 NDVI images were processed using ISODATA clustering algorithm (Mather and Koch, 2011), an unsupervised classification method, available in the Erdas Imagine software (Erdas, 2003). A series of classification runs corresponding to different number of clusters (2 to 30) were used. The ISODATA algorithm tries to minimize the Euclidian distance to form clusters. The results of the different runs are compared using the divergence separability which is a statistical measure of distance (Landgrebe, 2003); the ‘best’ number of clusters is the one corresponding to the run having the highest minimum and/or average divergence (Swain and Davis, 1978). The maximum number of iterations was 50 and the divergence threshold was 1. The spectral signatures of the clusters are represented graphically and similar NDVI profiles are merged to reduce the unnecessary large number of clusters.

Once the number of clusters is known, the NDVI profile clusters map is established and compared to the land cover map to match the preliminary legend of the former with the legend of the latter and also to get an idea about the land cover classes that are present in each of the NDVI profile clusters.

The NDVI profile clusters map, a raster, is converted to polygons and cropland areas are masked by using the land cover map to keep only NDVI profile clusters corresponding to agricultural land (Maselli and Rembold, 2001; Kastens et al, 2005).

Using GIS spatial analysis functions from ArcGIS (ESRI, 2009), the Mandals and the agricultural masked NDVI profile clusters map are overlaid to determine the respective areas (Ha) of each NDVI profile cluster by Mandal. These areas are further used as explanatory or independent variables, in the stepwise multiple linear regression (Neter and al, 1996), with the cropped areas (Ha) from agricultural statistics by season, crop, and Mandal as dependent variable:

$$CA = \sum_{i=1}^{n} c_i * NDVIcluster_i \qquad \text{Eq (3)}$$

with CA representing cropped area (Ha) by Mandal and $NDVIcluster_i$ representing the area (Ha) of the i^{th} NDVI profile cluster.

No constant was considered in the regression and the coefficients c_i were constrained to the 0 – 1 range in order to determine the estimated fraction or percentage of total area of a given NDVI profile cluster where a

given crop was grown at a given Mandal and a given season. Once the regression coefficients were estimated, the above equation was used to generate maps showing cropped fractions by map units. Statistical computations were done using the SPSS software (SPSS, 2008).

Results and Discussion

Average and minimum divergence values between clusters corresponding to the different runs (2 to 30 clusters) are reported in Figure 1.

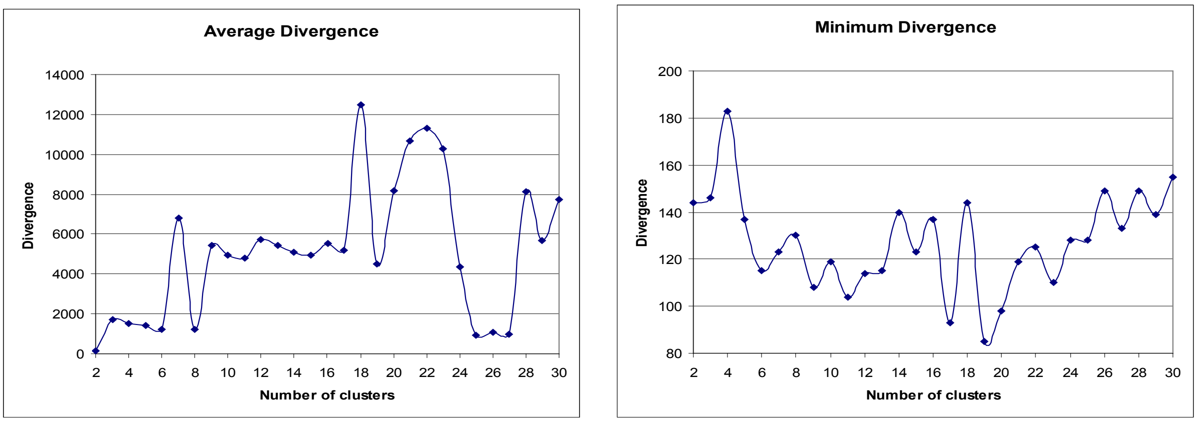

Figure 1. Average (left) and minimum (right) separability divergence values.

The highest value for the average divergence corresponds to 18 clusters whereas the highest one for minimum divergence corresponds to only 4 clusters while 18 clusters resulted in a reasonable high divergence value. So, based on these results, the optimal number of clusters given the best separability between them was taken to be 18. The corresponding average spectral signatures are displayed on Figure 2.

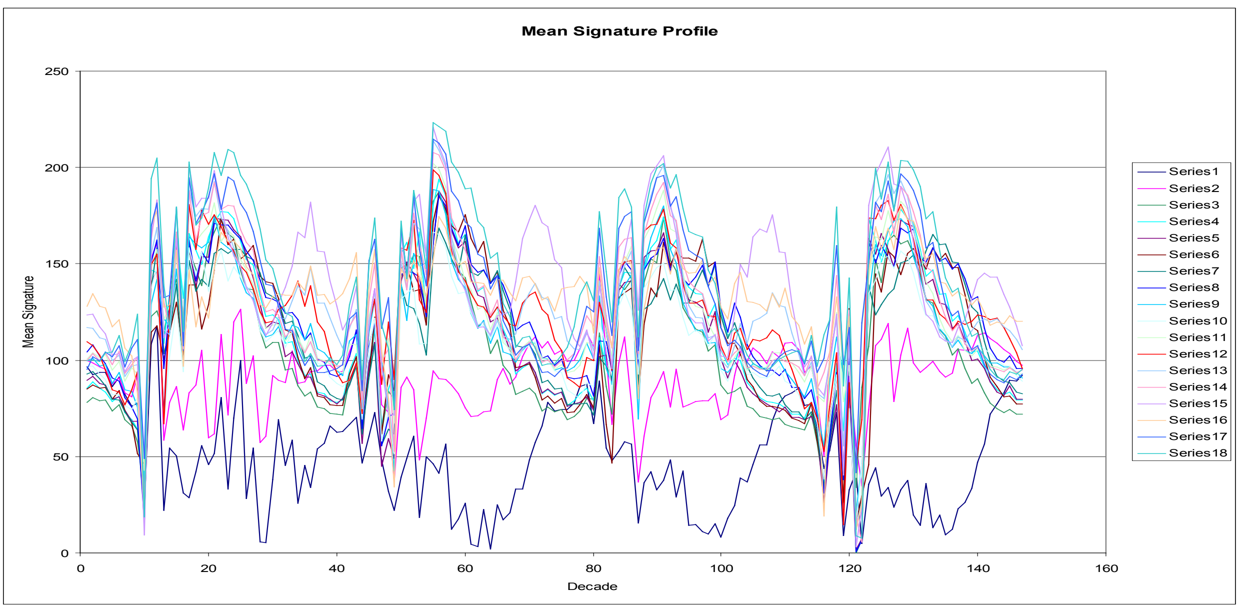

Figure 2. Average spectral signatures of the 18 NDVI profile clusters.

From this figure, some profiles (mainly 1, 2, and 15) have a distinct pattern while most of the others have

a more or less similar pattern. Most similar profiles were merged: profiles 3 to 7 and 10; profiles 8, 9, 11 and 14; profiles 12 and 13; and profiles 17 and 18. This combining of NDVI profiles resulted finally in only eight clusters (Figure 3) and the corresponding NDVI-unit map is displayed in Figure 4.

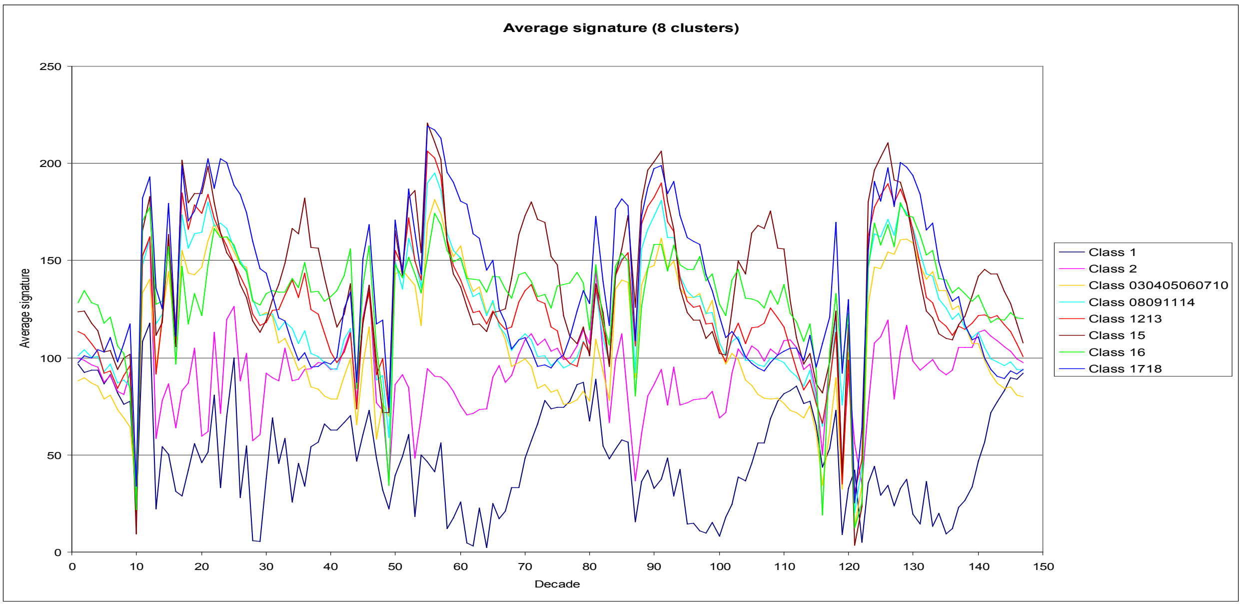

Figure 3. Average spectral signatures of remaining 8 NDVI profile clusters after merging similar ones.

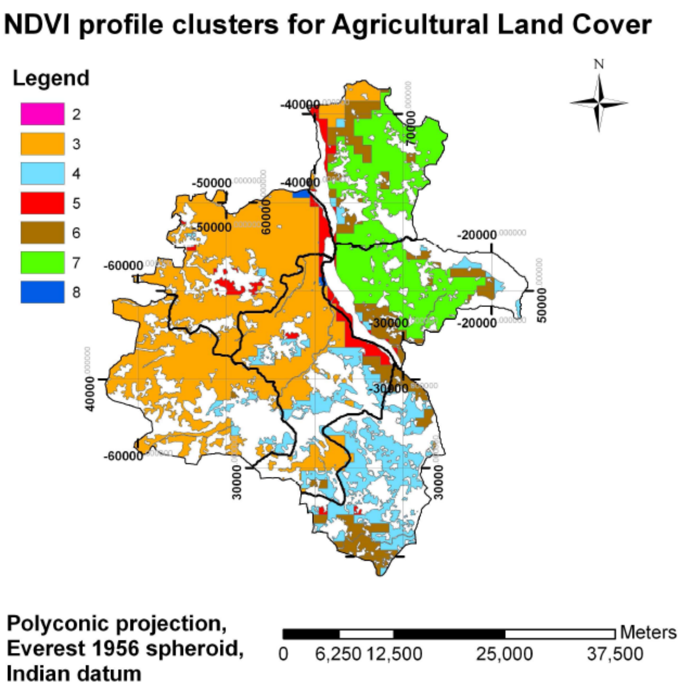

Figure 4. NDVI profile cluster map. N.B: cluster 1 is not present in the six Mandals.

White zones correspond to non agricultural areas. NDVI profile cluster 1 was present in the whole image of the Nizamabad district but not in the six Mandals or sub-districts corresponding to the study area. Also, clusters 2 and 8 are largely under-represented whereas clusters 3, 4, and 7 are much more present. These results

are confirmed by statistics provided in Table 1 which show the relation between agricultural land cover (either in Kharif, Rabi or both seasons) and the NDVI profile clusters and their corresponding areas and fractions. Cluster 3 is present in almost half of the total agricultural area, cluster 7 in fifth of this area and clusters 4 and 6 in 15 and 10 %, respectively. In contrast, clusters 2 and 8 are present in less than 1% while cluster 5 is present in 4% of the total agricultural area.

The results of stepwise multiple linear regression, for the main crops by season, are reported in Table 2.

This table shows that clusters 2, 5, and 8 are not involved in the regression models, at least, for the crops used at this step. This relates directly to the very limited extent of these clusters (see Table 1). Rice was grown, in the Kharif season, on 50% of the area of map-units represented by the NDVI-profile group 4 and 75% of the area of group 7 while it was grown, in Rabi season, on 2, 69 and 25% of areas of NDVI-profile groups 2, 6, and 7, respectively.

Table 1

Areas (Ha) and percentage of agricultural land cover by season corresponding to each NDVI-unit

NDVI unit / Season	Kharif	Rabi	Both seasons	Area (Ha)	Percentage
2	1	8	0	9	0.01
3	21602	19806	1001	42409	48.92
4	9875	2199	1414	13488	15.56
5	1289	1104	985	3378	3.90
6	3067	2426	3426	8920	10.29
7	240	3393	14583	18216	21.01
8	164	93	7	264	0.30
Total Area (Ha)	36239	29029	21417	86685	100
Percentage	41.81	33.49	24.71	100	

Table 2

Adjusted R^2 and coefficients (%) for stepwise multiple linear regression with total areas (Ha) for main crops in both seasons

		NDVI units				
Kharif	Adjusted R^2	3	4	6	7	Area (Ha)
Cotton	87.5	15.6				6860
Maize	81.3		4.1			482
Pulses	96.9	48.0	64.1			29121
Rice	95.0		50.3		75.3	22774
Sugarcane	89.9			26.0		2395
Rabi						
Groundnut	80.3			53.2		5942
Pulses	80.9	5.5				2824
Rice	99.8	1.8		69.1	25.0	11481
Sorghum	86.1	32.5				15454
Sugarcane	85.9			21.6		1960
Total Area (Ha) for both seasons		42409	13488	8920	18216	

The above regression coefficients were used to generate map of crops. For illustration, the map of rice, for both seasons, is displayed in Figure 5.

The comparison of these two maps shows that rice is cropped in both seasons mainly in 3 Mandals: the

southernmost one and the two located in the North-East of the study area. Regarding the southernmost Mandal, rice is not cropped in the same, but different, locations during the two seasons. For the two North-East Mandals, the non cropped areas (0%) in Kharif were intensively cropped (69%) in Rabi whereas the areas intensively cropped (75%) in Kharif were moderately cropped (25%) in Rabi.

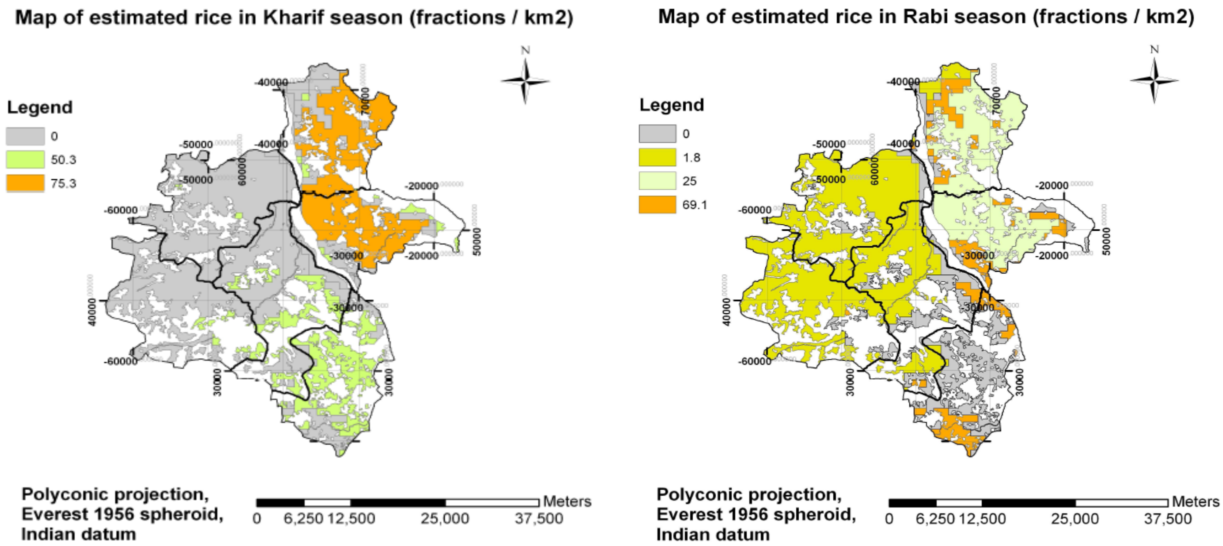

Figure 5. Estimated maps for rice grown in Kharif (left) and Rabi (right) seasons.

Conclusions

This research work illustrated the benefit of integrating hypertemporal remote sensing data with crop statistics to delineate NDVI profile clusters with their corresponding agricultural land cover map units and to link these statistics to geographical locations (mainly administrative units). These map units can be used as a reference for future monitoring of natural resources, in particular crop growth and development and consequently for forecasting crop production and/or yield and stresses like drought.

References

Ali A, De Bie CAJM, Skidmore AK, Scarrott RG, Hamad A, Venus V, and Lymberakis P. (2013). Mapping land cover gradients through analysis of hyper-temporal NDVI imagery. *International Journal of Applied Earth Observation and Geoinformation*, 23: 301–312

Ali A, De Bie CAJM, Skidmore AK, Scarrott RG, and Lymberakis P. (2014). Mapping the heterogeneity of natural and semi-natural landscapes. *International Journal of Applied Earth Observation and Geoinformation*, 26: 176–183

Allen JD. (1990). A look at the remote sensing applications program of the National Agricultural Statistics Service. *Journal of Official Statistics*, 6: 393-409

Atkinson PM, Jeganathan C, Dash J, and Atzberger C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of the Environment*, 123: 400–417

Balaghi R, Tychon B, Eerens H, and Jlibene M. (2008). Empirical regression models using NDVI, rainfall and temperature data for early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation*, 10: 438-452

Campbell NA, De Boer ES, and Hick PT. (1987). Some observations on crop profile modeling. *International Journal of Remote Sensing*, 8: 193-201

Carfagna E and Gallego FJ. (2005). Using remote sensing for agricultural statistics. *International Statistical Review*, 73: 389–404

Cihlar J. (2000). Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, 21: 1093 – 1114

CPO. (2001). Handbook of statistics – Nizamabad District, Chief Planning Officer: Nizamabad, Andhra Pradesh, India

Das DK, Mishra KK, and Kalra N. (1993). Assessing growth and yield of wheat using remotely sensed canopy temperature and spectral indices. *International Journal of Remote Sensing*, 14: 3081– 3092

ERDAS (2003). ERDAS field guide. 7th Ed. GIS and Mapping, LLC, Atlanta, Georgia and Leica Geosystems GIS and Mapping, LLC.

ESRI. (2009). ArcGIS: ArcMap 9.2. ESRI: Redlands, California, USA.

Falkner E and Dennis M. (2002). Aerial mapping: methods and applications. 2nd Edition, CRC Press: Boca Raton, Florida, USA

Ferencz C, Bognar P, Lichtenberger J, Hamar D, Tarcsai G, and Timar G. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, 20: 4113–4149

Giri CP. (2012). Remote sensing of land use and land cover: Principles and applications. CRC Press: Boca Raton, Florida, USA

Gonzalez-Alonso F, Cuevas JM, Arbiol R, and Baulies X. (1997). Remote sensing and agricultural statistics: Crop area estimation in north-eastern Spain through diachronic Landsat TM and ground sample data. *International Journal of Remote Sensing*, 18: 467-470

Howard DM, Whlie BK, and Tieszen LL. (2012). Crop classification modeling using remote sensing and environmental data in the Greater Platte River Basin, USA. *International Journal of Remote Sensing*, 33: 6094-6108

Jiang Y, De Bie CAJM, Wang T, Skidmore AK, Liu X, Song S, and Shao X. (2013). Hyper-temporal remote sensing helps in relating epiphyllous liverworts and evergreen forests. *Journal of Vegetation Science* 24 (2013) 214–226

Kastens JH, Kastens TL, Kastens DLA, Price KP, Martinko EA, and Lee RY. (2005). Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sensing of Environment*, 99: 341 – 356

Khan MR, De Bie CAJM, Van Keulen H, Smaling EMA, and Real R. (2010). Disaggregating and mapping crop statistics using hypertemporal remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 12: 36–46

Knight JF, Lunetta RS, Ediriwickrema J, and Khorram S. (2006) Regional scale land cover characterization using MODIS-NDVI 250m multi-temporal imagery: a phenology-based approach. *GIScience and Remote Sensing*, 43: 1–23

Labus MP, Nielsen GA, Lawrence LR, Engel R, and Long DS. (2002). Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing*, 23: 4169-4180

Landgrebe DA. (2003). Signal theory methods in multispectral remote sensing. Wiley: Hoboken, New Jersey, USA

Lillesand T, Kiefer RW, and Chipman J. (2007). Remote sensing and image interpretation. 6th Ed. Wiley: New York, USA

Maselli F and Rembold F. (2001). Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries. *Photogrammetric Engineering and Remote Sensing*, 67: 593– 602

Mather PM and Koch M. (2011). Computer processing of remotely-sensed images: an introduction. 4th ed. Wiley-Blackwell: Chichester, UK

Mkhabela MS, Mkhabela MS, and Mashinini NN. (2005). Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. *Agricultural and Forest Meteorology*, 129: 1–9

Murakami T, Ogawa S, Ishitsuka N, Kumagai K, and Saito G. (2001). Crop discrimination with multi-temporal SPOT/HRV data in the Saga Plains, Japan. *International Journal of Remote sensing*, 22: 1335-1348

Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W. (1996). Applied linear statistical models. 4th ed. McGraw-Hill: Boston, Massachusetts, USA.

Nguyen TTH, De Bie CAJM, Ali A, Smaling MA, and Chu TH. (2012). Mapping the irrigated rice cropping patterns of the Mekong delta, Vietnam, through hyper-temporal SPOT NDVI image analysis. *International Journal of Remote sensing*, 33: 415-434

Nordberg ML and Evertson J. (2003). Vegetation index differencing and linear regression for change detection in a Swedish mountain range using Landsat TM and ETM+ imagery. *Land Degradation and Development*, 16:139–149

Philipson W. (1997). Manual of photographic interpretation. 2nd Ed. Bethesda. MD: American Society for Photogrammetry and Remote Sensing

Potgieter AB, Lawson K, and Huete AR. (2013). Determining crop acreage estimates for specific winter crops using shape attributes from sequential MODIS imagery. *International Journal of Applied Earth Observation and Geoinformation*, 23: 254–263

Rouse JW, Haas RH, Schell JA, and Deering DW. (1973). Monitoring vegetation systems in the Great Plains with ERTS. Third Earth Resources Technology Satelite-1 Symposium (pp. 309-317). Washington, D.C.

Sakamoto T, Yokozawa M, Toritani H, Shibayama M, Ishitsuka N, and Ohno H. (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of the Environment*, 96: 366–374

Sarkar S and Kafatos M. (2004). Inter-annual variability of vegetation over India sub-continent and its relation to the different meteorological parameters. *Remote sensing of the Environment*, 90: 268-280.

Sheub SS and Atkins PJ. (1991). Crop area estimation: a comparison of remote sensing and census methods. *Geography*, 76: 235-239

SPSS. (2008). SPSS Statistics for Windows, Version 17.0. SPSS: Chicago, Illinois, USA.

Swain PH and Davis SM. (1978). Remote sensing: the quantitative approach. McGraw-Hill: New York, USA

Tarpley J, Schneider S, and Money R. (1984). Global vegetation indices from NOAA-7 meteorological satellite. *Journal of Climate and Applied Meteorology*, 23: 491–494

Taylor JC, Brewer TR, and Bird AC. (2000). Monitoring landscape changes in the National Parks of England and Wales using aerial photo-interpretation and GIS. *International Journal of Remote Sensing*, 21: 2737-2752

Tucker CJ. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of the Environment*, 8: 127-150

USDA. (2014). Crop production 2013 summary. United States Department of Agriculture: National Agricultural Statistics Service

White MA, De Beurs KM, Didan K, Inouye DW, Richardson AD, Jensen OP, O'Keefe J, Zhang G, Nemani RR, and Van Leeuwen WJD. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology*, 15: 2335–2359

Xie Y, Sha Z, and Yu M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1: 9-23

You X, Meng J, Zhang M, and Dong T. (2013). Remote sensing based detection of crop phenology for agricultural zones in China using a new threshold method. *Remote Sensing*, 5 : 3190-3211

Journal of Statistical Science and Application 2 (2014) 102-110

Crop Yield Estimation with Farmers' Appraisal on Weather Condition

Jean Baptiste HABYARIMANA
Agriculture Statistician
Ministry of Agriculture and Animal Resources
Kigali, Rwanda

Crop yield is mainly affected by weather condition, inputs, and agriculture policies. In the crop yield estimation, farmers' perception on weather conditions lead to the assessment of how well yield would be compared to the previous seasons. This paper applies Bayesian estimation method to estimate crop yield with farmers' appraisal on weather condition. The paper shows that crop yield estimation with farmers' appraisal on weather condition takes into account risk proportionally to climate change. In light of the United Nations efforts aimed to build a consolidated agriculture statistical system across countries, the statistical model developed here should provide an important tool both for the crop yield estimation and food price analysis.

Key words: Crop Yield Estimation, Farmers' Appraisal on Weather Condition, Crop growing condition, Bayesian Estimation Method

Introduction

Land productivity is a vital factor in feeding the population. It is also a critical factor in the struggle of developing countries to improve the availability of food. In line with MDGs, developing countries have undertaken different measures to improve land productivity. Following all those policies the main challenge for agriculture statistician in many developing countries remains to be the development of statistical models that could provide reliable crop yield estimates with high exactitude to monitor and assess the progress of agricultural land productivity. This study undertakes this task to offer a method complementary to those available in the literature such as crop cutting and famers' estimate.

Literature Review

In most of the cases, yield forecasts are based on reports by crop correspondents at regular intervals during the growing season using crop appearance as an indicator. In this context, the most accurate method of estimation consists of crop cutting method" (G.R. Spinks).

Thus, crop condition data has the potential to provide a simple, regular source of information about the realized yield. In addition, weather affects crops differently during different stages of crop growth (Ranjana Agrawal, 2012).

Corresponding author: Jean Baptiste HABYARIMANA, Agriculture Statistician, Ministry of Agriculture and Animal Resources, Kigali, Rwanda. E-mail: kradrimn@gmail.com.

To overlook the impact of weather condition on crops growing and yields, Stasny, Goel and other researchers at the Ohio State University developed a Bayesian mixed-effects county yield estimation algorithm with a spatial component involving correlations among neighboring counties (Michael E. Bellow, 2007). In addition to this, Bayesian probability approach of obtaining yield forecasting involves the collection of expert opinion data of farmers who are actually engaged in raising the crop regarding their assessments about the likely crop production (Chandrahas).

Estimation Method

Drawing on the crop yield forecasting and estimation literature, crop yield is a function of weather condition, inputs and agro-ecological conditions including weather, land and external input use. To cover the gap between crop yield and risk resulting from climate change, this paper models crop yield as a function of weather and crop growing conditions (appearance) subjected to farmers' information on weather.

Statistical Model

From a Bayesian standpoint, true model parameters were unknown and therefore considered to be random, and they were assigned a joint probability distribution. Prior distribution was used to summarize our state of knowledge, or what is currently known about the parameters. Once the data were observed, the evidence provided by the data was combined with the prior distribution using Bayes' Theorem. The result of combining prior information and empirical evidence was an updated posterior distribution for the parameters.

In this context, four piece of information was used "actual yield, yield targets, realized performance towards yields targets, and farmers' appraisal on weather condition". Those pieces of information were also used to estimate variance and covariance information on yield. In the empirical example for model illustration six crops were considered when trying to forecast yields.

In the first stage of parameter estimation, average yields "y_{fj}" was computed using actual yields of Maize, Rice, Beans, Cassava, Irish Potatoes, and farmers' appraisal on weather conditions:

Box 1

$$y_{fj} = \frac{\sum_{i=1}^{4} y_{ij}\,\omega_{ij}\,c_{ik}}{\sum_{i=1}^{4} \omega_{ij}\,c_{ik}}$$

Where $\omega_{ij} = y_{ij} * p_{ij}$ and $p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{4} y_{ij}}$

Where y_{ij} = {1, 2, 3, 4 | yield values for four percentiles dividing actual yields into five equal groups} for crop j = Maize, Rice, Wheat, Beans, Cassava, and Irish Potatoes. Four scales weather conditions are denoted by k = Poor, Fair, Good, Very Good; and ω_{ij} = expected yield in each scale for crop "j". Weather conditions "C_{ik}" are associated with actual yield y_{ij} by C_{ik} with {C_{ik} = {1=Poor, 2=Fair, 3=Good, 4=Very Good | i percentile is linked with its corresponding k weather condition}.

In the second stage, actual performance to reach yield targets for each crop "j" "R_j" was computed using average yields "y_{fj}" and yield targets "y_{tj}"

Box 2

$$R_j = \frac{y_{fj}}{y_{tj}}$$

In the third stage, average yields "y_{fj}", yield targets Y_{tj} and actual yield performance R_j and their associated reliabilities were used to estimate covariance and variance information on yields.

To develop Bayesian methods for generating yield estimates from readily available crop and weather conditions information; it was assumed that average yield are normally distributed “$d_1|x^1 \sim N_p(D_1x^1\ \Sigma_1)$” and that posterior distribution of yields is given by $x^1|d_1 \sim N_n(\mu;\ V)$” where μ_{t+1} = Estimated average yield and V = Variance.

$\mu = (A^{+}HA^{+T})D^{T}{}_{1}\Sigma^{-1}d1)$ and

$V = A^{+}HA^{+\mathrm{T}} - A^{+}HA^{+T}D^{T}{}_{1}\Sigma^{-1}D1A^{+}HA^{+T}$

$A^{+}=A^{T}(AA^{T})^{-1}$ "Moore-Penrose inverse"

and $\Sigma_0 = \Sigma_1 + D1A^{+}HA^{+T}D^{T}$.

(see Theorem 1 in Magnus, Tongeren and Vos (2000).

Case study Development

Agriculture sector is predominant in Rwanda and the Government of Rwanda has invested a lot in agriculture sector by introducing new seed varieties, reinforcing the use of chemical fertilizers and pesticides, soil management and land rehabilitation, anti-erosion activities, farmers field school, etc.. But controversially, Rwanda experiences food prices fluctuation over time. Hence, as growing season forecasts of crop yields are of considerable interest to commodity market participants and price analysts, the main problematic issue remains to be the development of consistent method with compatible predicting statistical model able to make yield estimations with much higher precision to explain those controversial phenomena.

The yields data for empirical illustration used in this paper came from Crop Assessment Surveys.

This paper hypotheses that Bayesian estimation method that consider past information on crop yields, present farmers’ knowledge on weather conditions and present agro-ecological information could provide a simpler and peerless complementary statistical model to estimate crop yield.

Results and Discussion

Farmers’ problem is an optimization problem in which estimated yield should be close to forecasted yields as possible. The optimization problem shows that:

(1) The expected yields are less than average of yields recorded in 22 past seasons;

(2) Forecasted yields are less than the average of yields of two latest previous seasons;

(3) Estimated yields are greater than average of yields recoded in 22 past seasons.

(4) The final results show that the estimated crop yields tend to deviate from the forecasted crop yields by the mean of yields recorded in 22 past seasons.

Descriptive Statistics

Table 1 shows descriptive statistics “Mean yield (Kg/Ha), Standard deviation, Minimum and maximum yields (Kg/Ha) realized in 11 past years with 22 observations”.

Figure 1, shows the trends (2002 to 2012) of Maize, Wheat, Rice, Beans, Irish Potato and Cassava yields. Trends show an overtime increase of yields for all crops considered.

Crop Yield Estimates

The final results showed in table 2 revealed that when estimated yields for 2013 Agriculture year are compared to actual yields realized in 2012 Agriculture Season; yield (Kg/Ha) for Maize, Wheat, Rice, Irish Potato, and Cassava and Beans could increase if the weather condition does not change or change slightly as it was appraised by farmers.

Table 2 shows that expected yield are less than the mean of actual yields. The mean of actual yields are less than forecasted yields. Forecasted yield are less than estimated yields.

Table 1

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
maize	22	1264	717	559	2820
wheat	22	1096	598	377	2208
rice	22	4020	1063	1890	5751
beans	22	837	190	500	1101
iris potato	22	8003	2241	5409	12605
cassava	22	9093	3135	5177	13974

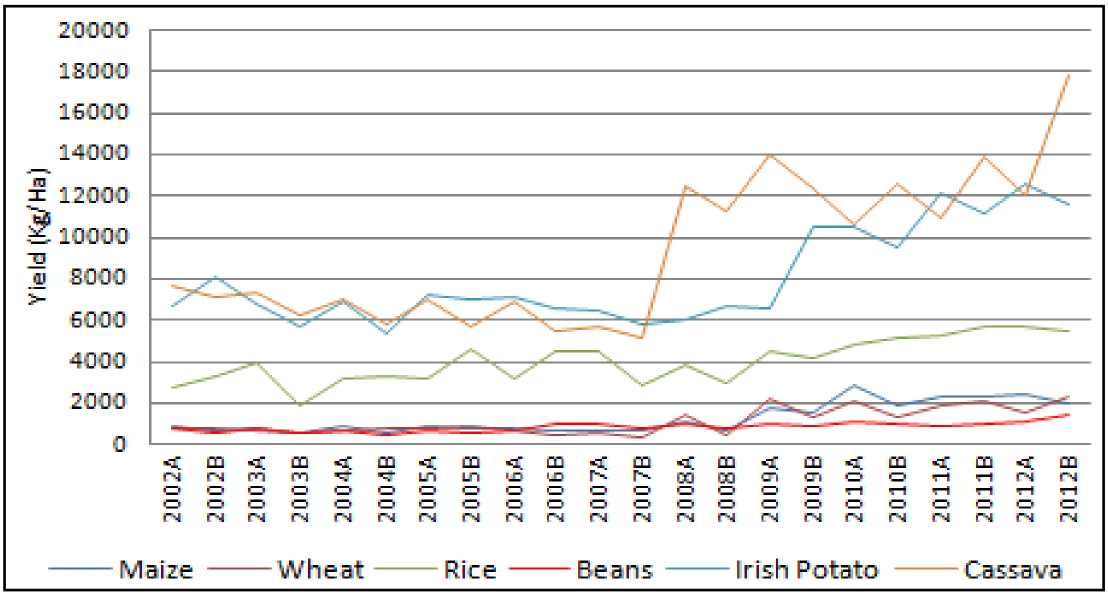

Figure 1. Trend of Crop Yield over last 11 Years.

Table 2

Crop Yields Estimates

	Maize	Wheat	Rice	Beans	Irish Potato	Cassava
Y_j = Actual Yields (2002 – 2012)	1,260	1,125	4,031	855	8,055	9,326
$\hat{Y}_j$ = Expected Yields	1,123	1,118	3,378	701	6,497	7,847
Y_{2012} = Actual Yields	2,186	1,916	5,597	1,265	12,115	14,934
Y_{tj} = Targeted Yields[1]	3,750	3,500	7,000	1,850	27,500	35,000
Y_{fj} = Forecasted Yields	1,459	1,440	3,980	829	7,809	9,417
Y_{ej} = Estimated Yields for 2013A	2,934	2,889	5,878	1,397	17,793	22,965

Assessment of the progress of agricultural land productivity

To assess the progress of land productivity to feed population and make available food at markets, performances were computed as the basis of crop yield monitoring. i) Performance with forecasted yields that

[1] Yields collected in Rwanda Agriculture Agenda "Agenda Agricole"

combines past information, states what was the performance to reach targeted yields by the time of setting those targets; ii) Actual performance was computed referring to the average yields of the latest two agriculture seasons to state how far agriculture households were to reach targeted yields in 2012; iii) Performance with estimated yields was computed referring to the estimated yields and targeted yields to state how far agriculture households should be to reach targeted yields in 2013.

Table 3

Land productivity assessment

	Maize	Wheat	Rice	Beans	Irish Potato	Cassava
R_j = Performance with Forecasted Yield	39%	41%	57%	45%	28%	27%
Actual Performance	67%	75%	71%	66%	64%	63%
Performance with Estimated Yield	78%	83%	84%	76%	65%	66%

Performance ratios derived from the model formulated using prior information and empirical evidences could in meaningful way be used in land productivity and food market price analyses. Performance ratios could also describe the relationship between agriculture production on one hand and the effort made to improve land productivity such us introduction of new seeds varieties; use of fertilizers and pesticides; soil management and extension services on the other hand. The used model could play an important role when analyzing the impact of climate change on agriculture sector and the effort made to face the problem of climate change.

Policy Implication

The estimation model developed in this paper could help decision takers and policymakers: i) To monitor land productivity and yield targets; ii) To link crop yield and crop production; iii) To monitor food availability, food demand and market access when assessing food shortage and planning for food redistribution"; iv) To link weather condition a constraints to land productivity and food availability with meaningful early warnings to lower crop production risks and ensure public awareness and preparedness to act; v) To correlate food prices with agriculture production and climate change.

Concluding Remarks

This paper develops predicting statistical model laying on Bayesian Method. The developed Statistical model for crop yield estimation could contribute to the development and credibility of agricultural statistics. This paper comes with new insights and paves the way for agricultural policy analysis by providing timely high precision, credible and compatible yield estimates that could lead to reliable crop production estimation. As crop yields have significant impact for both commodity prices and farmer income, growing season estimates of crop yields provided by this model could be of considerable utilization to price and food security analysts.

Acknowledgements

I wish to express my profound gratitude to Dr. Temel Tugrul senior researcher, Mr. Enock Chinganda senior expert in statistics, Birasa Nyamulinda lecturer at National University of Rwanda, and Mrs Francine UWAYO agricultural statistician at MINAGRI who provided technical advices. Their comments and guidance were very useful to the study. The errors in this paper, however, are the author's and neither Government Institutions nor Non-Government Institutions is responsible for them.

References

Alpha C. (1984) Chiang. *Fundamental Methods of Mathematical Economics*. Third Edition. USA.

Carl P Simon and Lawrence Blume. (1994) *Mathematics for Economists.* First Edition. USA.

Chandrahas, *Bayesian Probability Approach for Crop Yield Forecasting*. Indian Agricultural Statistics Research Institute, New Delhi-110012).

Fryzlewicz, P., Van Bellegem, S. & von Sachs. (2003) *Forecasting non-stationary time series by wavelet process modelling*. Annals of the Institute of Statistical Mathematics.

G.R. Spinks. *Uses and Methods of Crop Forecastin.*

Kruse, John R. and Darnell Smith. "*Yield Estimation Throughout the Growing Season*." In Applied Price Analysis, Forecasting, and Market Risk Management, Proceedings of the NCR-134.

Magnus, J. R., van Tongeren, J., and Vos. (2000) National accounts estimation using indicator ratios. Review of Income and Wealth, Series 46, No. 3, p-329-350.

Michael E. Bellow. (2007) *Comparison of Methods for Estimating Crop Yield at the County Level.*

Paul L. Fackler and Bailey Norwood, *Forecasting Crop Yields and Condition Indices*.

R. Davidson and J.G. MacKinnon. (1993) *Estimation and Inference in Econometrics*. Oxford University Press, Oxford.

Ranjana Agrawal. (2012) *Forecasting Techniques in Crops*. New Delhi.

Vladimir Rakocevic, On continuity of the moore-penrose and drazin inverses, Math. P: 163-172

USDA. (1998) "*Crop Progress/Crop Weather: Terms and Definitions*." National Agricultural Statistics Service {NASS), Washington.

USDA. *Crop Production*. National Agricultural Statistics Service (NASS), Washington. Monthly.

Annexes

Methodology

To perform Bayesian Estimation Method for Crop yield forecasting, five matrixes d_{6x1} Matrix of Targeted yields, $[D_{6x12} = (I_{6x6}\ 0_{6x6})]$ Matrix for Crops to be used in yield forecasting, A_{12x12} Matrix of Actual level to reach targeted Yield "Y_{tj}", Variance Matrix (E_1) , Covariance Matrix (H) were estimated and used in the Bayesian Estimation Model. The matrices dimensions follow the number of crop which we want to estimate their yield. In the empirical example for model illustration six crops were considered which yield estimates are needed for each and every crop (6+6=12). The full steps to estimates those matrices are illustrated bellow:

In the first stage of estimating parameters, Forecasted Average yield "y_{fj}" using actual yields for Maize, Rice, Beans, Cassava, Irish Potatoes, and Farmers Appraisal on weather condition was computed as follow:

$$y_{fj} = \frac{\sum_{i=1}^{4} y_{ij}\, \omega_{ij}\, c_{ik}}{\sum_{i=1}^{4} \omega_{ij}\, c_{ik}}$$

Where $\omega_{ij} = y_{ij} * p_{ij}$ with $p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{4} y_{ij}}$

Where y_i = {1, 2, 3, 4 | yield values for four percentiles dividing actual yields into five equal groups} for crop "j" with j = {Crops | Maize, Rice, Wheat, Beans, Cassava, and Irish Potatoes}, k = {Poor, Fair, Good, Very Good | C_k=Weather Condition}; and ω_{ij} = expected yield in each category for crop "j". Weather Conditions "C_{ik}" are associated with Actual yield y_{ij} by C_{ik} with {$C_{ik} =$ {1=Poor, 2=Fair, 3=Good, 4=Very Good | i percentile is linked with its corresponding k weather condition}.

In the second stage, actual performance to reach yield targets for each crop "j" "R_j" is computed basing on y_{fj} = Forecasted Yield for Crop (j) and Y_{tj} = Yield Target for (j):

$$R_j = \frac{y_{fj}}{y_{tj}}$$

In the third stage, Y_{fj} and their associated reliabilities (in this paper reliability was assigned to be High Medium "HM = 3%"

and Medium "M = 6%") were used to estimate Variance Matrix (E_1); while Y_{tj}, R_j, and their associated reliabilities were used to estimate Covariance Matrix (H). Matrix D was estimated using Identity Matrix (6x6) of 6 Six crops used in this paper and Zero Matrix (6x6) of forecasted Yield for six crop therefore $D_{6x12} = (I_{6x6}\, 0_{6x6})$, the column matrix ($d_{6x1}$) of Laboratory yield.

Assuming that "$d_1|x^1 \sim N_p(D_1x^1\ \Sigma_1)$ where $D_{1,(p;n)}$ has full row-rank and Σ_1 is positive definite (hence non-singular); (ii) $Ax^1 \sim N_m(h;H)$ where $A = (A1 : A2)$, a column vector $h = (h_1; h_2)$; a block diagonal matrix $H = (H1;H2)$ with H_1 associated with A_1 and H_2 with A_2; (iii) A has full row- rank and H may be singular. Then the posterior distribution of x^1 is given by $x^1|d_1 \sim N_n(\mu; V)$" with:

$$\mu = (A^+HA^{+T})D^T{}_1\Sigma^{-1}d_1)$$

and

$$V = A^+HA^{+T} - A^+HA^{+T}D^T{}_1\Sigma^{-1}D1A^+HA^{+T}$$

Where $A^+ = A^T (AA^T)^{-1}$ "the Moore-Penrose inverse", and

$$\Sigma_0 = \Sigma_1 + D1A^+HA^{+T}D^T$$

Description of data

Crop Yields 2002 – 2012 (22 Agricultural Seasons)

Yield (Kg/Ha)	Maize	Wheat	Rice	Beans	Irish Potato	Cassava
2002A	927	800	2,714	755	6,682	7,700
2002B	664	755	3,335	595	8,067	7,098
2003A	804	725	4,000	773	6,818	7,318
2003B	577	538	1,890	550	5,682	6,255
2004A	845	725	3,188	709	6,864	7,045
2004B	559	760	3,300	500	5,409	5,773
2005A	860	785	3,188	709	7,259	7,032
2005B	736	940	4,583	514	7,000	5,727
2006A	824	628	3,188	680	7,168	6,905
2006B	709	503	4,525	1,007	6,535	5,478
2007A	723	533	4,525	988	6,500	5,671
2007B	706	377	2,817	833	5,800	5,177
2008A	1,114	1,430	3,791	988	6,050	12,425
2008B	718	419	2,917	831	6,667	11,300
2009A	1,797	2,208	4,479	1,019	6,533	13,974
2009B	1,551	1,371	4,159	843	10,537	12,345
2010A	2,820	2,127	4,786	1,076	10,563	10,616
2010B	1,853	1,329	5,137	970	9,563	12,609
2011A	2,270	1,924	5,211	937	12,102	10,917
2011B	2,283	2,039	5,751	1,011	11,186	13,933
2012A	2,406	1,562	5,725	1,101	12,605	12,072
2012B	1,965	2,270	5,469	1,428	11,625	17,795

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
maize	22	1264	717	559	2820
wheat	22	1096	598	377	2208
rice	22	4020	1063	1890	5751
beans	22	837	190	500	1101
iris potato	22	8003	2241	5409	12605
cassava	22	9093	3135	5177	13974

	Cut-of	Y_{ij}	p_{ij}	ω_{ij}	C_{ik}	$Y_{ij}\omega_{ij}C_{ik}$	$\omega_{ij}C_{ik}$
Maize Percentiles	20	708	0.15	107	0.476	36,077	51
	40	808	0.17	139	0.247	27,818	34
	60	1077	0.23	248	0.148	39,489	37
	80	2087	0.45	931	0.128	248,604	119
Wheat Percentiles	20	536	0.12	64	0.476	16,201	30
	40	756	0.17	127	0.247	23,639	31
	60	1252	0.28	347	0.148	64,283	51
	80	1970	0.44	860	0.128	216,844	110
Rice Percentiles	20	3079	0.19	586	0.476	858,479	279
	40	3426	0.21	725	0.247	613,516	179
	60	4516	0.28	1260	0.148	841,942	186
	80	5167	0.32	1649	0.128	1,090,556	211
Beans Percentiles	20	646	0.19	123	0.476	37,696	58
	40	784	0.23	181	0.247	34,970	45
	60	963	0.28	272	0.148	38,816	40
	80	1014	0.30	302	0.128	39,194	39
Irish Potato Percentiles	20	6320	0.20	1285	0.476	3,865,861	612
	40	6709	0.22	1448	0.247	2,399,805	358
	60	7241	0.23	1687	0.148	1,807,711	250
	80	10812	0.35	3761	0.128	5,205,257	481
Cassava Percentiles	20	5755	0.16	916	0.476	2,508,073	436
	40	7056	0.20	1377	0.247	2,399,246	340
	60	10857	0.30	3259	0.148	5,236,726	482
	80	12499	0.35	4319	0.128	6,910,351	553

Model Estimation

Estimation of Matrix D

Maize	Rice	Wheat	Beans	Cassava	Irish Potatoes	MaizeF	RiceF	WheatF	BeansF	CassavaF	Irish PotatoesF
1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0

Estimation of Variance Matrix (E Matrix)

Crops	Mean	Reliability	PrioSev= Mean * Re	PrioV = PrioSe^2
Maize	1260	0.01	13	159
Wheat	1125	0.05	56	3164
Rice	4031	0.05	202	40622
Beans	855	0.05	43	1828
Irish Potato	8055	0.05	403	162208
Cassava	9326	0.05	466	217436

Estimated Variance Matrix

Maize	Wheat	Rice	Beans	Irish Potato	Cassava
159	0	0	0	0	0
0	3164	0	0	0	0
0	0	40622	0	0	0
0	0	0	1828	0	0
0	0	0	0	162208	0
0	0	0	0	0	217436

Estimation of covariance Matrix (H Matrix)

	Yfj	Ytj	Pi	Reability	PrioSe	PrioV	Bij^2	PrioV*Bij^2
Maize	1,459	3750	0.39	0.01	0.004	0.0000	14062500	213
Wheat	1,440	3500	0.41	0.05	0.021	0.0004	12250000	5181
Rice	3,980	7000	0.57	0.05	0.028	0.0008	49000000	39600
Beans	829	1850	0.45	0.05	0.022	0.0005	3422500	1716
Irish Potato	7,809	27500	0.28	0.05	0.014	0.0002	756250000	152446
Cassava	9,417	35000	0.27	0.05	0.013	0.0002	1225000000	221679

Estimated Covariance Matrix

Maize	Wheat	Rice	Beans	Irish Potato	Cassava	MaizeF	WheatF	RiceF	BeansF	Irish PotatoF	CassavaF
213	0	0	0	0	0	0	0	0	0	0	0
0	5181	0	0	0	0	0	0	0	0	0	0
0	0	39600	0	0	0	0	0	0	0	0	0
0	0	0	1716	0	0	0	0	0	0	0	0
0	0	0	0	152446	0	0	0	0	0	0	0
0	0	0	0	0	221679	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Estmation of A Matrix (Pi)

Maize	Rice	Wheat	Beans	Cassava	Irish Potatoes	MaizeF	RiceF	WheatF	BeansF	CassavaF	Irish PotatoesF
-0.39	0	0	0	0	0	1	0	0	0	0	0
0	-0.41	0	0	0	0	0	1	0	0	0	0
0	0	-0.57	0	0	0	0	0	1	0	0	0
0	0	0	-0.45	0	0	0	0	0	1	0	0
0	0	0	0	-0.28	0	0	0	0	0	1	0
0	0	0	0	0	-0.27	0	0	0	0	0	1
-1	0	0	0	0	0	1	0	0	0	0	0
0	-1	0	0	0	0	0	1	0	0	0	0
0	0	-1	0	0	0	0	0	1	0	0	0
0	0	0	-1	0	0	0	0	0	1	0	0
0	0	0	0	-1	0	0	0	0	0	1	0
0	0	0	0	0	-1	0	0	0	0	0	1

Estimation of dT Matrix

	Maize	Rice	Wheat	Beans	Cassava	Irish Potatoes
Y_{fj}	1,459	1,440	3,980	829	7,809	9,417

Journal of Statistical Science and Application 2 (2014) 111-118

No More Excuses - Personality Traits and Academic Dishonesty in Online Courses

Yovav Eshet
The Department of Multi-Disciplinary Studies
Zefat Academic College
Zefat, Israel

Keren Grinautski
The Department of Multi-Disciplinary Studies
The Western Galilee College
Akko, Israel

Yehuda Peled
Department of Information Science
The Western Galilee College
Akko, Israel

Casimir Barczyk
School of Management
Purdue University, Calumet, USA.

Academic dishonesty is a disturbing issue in higher education that has been worsening over the years, especially with the appearance of the internet and the e-learning education. This new technology exposes students to the opportunity of using online bank exams and term papers and increases their tendency to cheat. This study investigates student academic dishonesty in the context of traditional and distance-learning courses in higher education. Data from 1,365 students enrolled in academic institutes in the U.S.A and Israel were surveyed to assess their personality and their willingness to commit various acts of academic misconduct. The findings indicate that in both countries dishonest behaviors are greater in face-to-face courses than in online courses. In addition, both American and Israeli students identified with the personality trait of Agreeableness showed a negative correlation with academic dishonesty. Furthermore, Israeli students identified with the personality traits of Conscientiousness and Emotional Stability demonstrated a negative correlation with academic dishonesty. In contrast, the personality trait of Extraversion among American students was found in a positive correlation with academic misconduct. Implications for further research are discussed.

Keywords: Academic Dishonesty, Personality Traits, OCEAN, Online Courses.

Corresponding author: Yovav Eshet, The Department of Multi-Disciplinary Studies, Zefat Academic College, Zefat, Israel.

Introduction

Academic dishonesty has been described as an act of cheating, deception and violation of rules for a personal gain or advantage [1], [2] done by the student, "a conscious effort to use proscribed data and/or resources on exams or written work submitted for academic credit" [3]. Before engaging in unethical academic behavior, a student has to make a rational decision that the benefit of cheating worth the risk of getting caught [4]. There are two types of academic dishonesty - active and passive, both including an intention for cheating. Active includes an act for raising a student's grade, whereas passive includes a behavior of assisting other student to raise his grade [5].

Researchers have shown that academic dishonesty has been worsened over the years [6], [7], [4], [8], and that cheating is an epidemic phenomena across most college campuses [9], [10], [11]. The development of information technology and the accessibility of academic material on the internet made it easier to engage in cheating and in plagiarism [12], [13], [14], and [15]. Furthermore, assignments and papers are available for purchase to students who seek for it [16].

In addition, the growth of technology encouraged the existence of online courses and distant education. According to the National Center for Education Statistics [17], almost 4.3 million undergraduate students are participating in online courses per year. There is a notion that it is easier to cheat when participating in distance learning classes [18], and both students and faculty are aware of the intensity of this phenomenon compared to traditional courses [11], particularly where there is little or no personal contact between students and faculty [19], [20]. Similarly, Kelley and Bonner [21] suggested that students who feel close to their professors tend to be more honest. However, in the online learning environment the ability for faculty to develop a strong rapport with students becomes more difficult. Students who feel “distant” from others seem to have higher tendency to perform deceptive behaviors, such as cheating [22], [23]. Online courses, in contrast to traditional classroom courses, may serve to exacerbate these feelings of separation and, thus, may contribute to higher incidence of academic dishonesty [24], [25].

The research literature shows that academic dishonesty is influenced not only by situational factors (circumstantial and contextual) [26], such as the teaching method – on-line vs. traditional classrooms, as we mentioned above, but individual factors as well (demographic, psychosocial and academic characteristics of students) [27]. One of these individual factors is related to various personality traits.

Students' Personality as a Predictor of Academic Dishonesty

Research regarding the relationship between unethical academic behavior and personality traits includes several studies, while each study uses a different measure of academic dishonesty. Hence, the results are often contradicting [28], [29] and [30]. Although the ability of the 'Big Five' personality traits measure was proved effective in explaining unethical behaviors [31], and would be expected to have a direct impact on the level of students' cheating behavior [32], it is not frequently used in the context of academic dishonesty and most researches who did use it addressed only few traits instead of the whole model [30], [31]. Below are explained the personality traits of the "Big Five" model in the context of academic dishonesty.

Big Five personality traits include Extraversion, Conscientiousness, Agreeableness, Neuroticism and Openness to experience [33]. The personality trait of Extraversion is characterized as the tendency to be talkative, assertive, energetic, sensation-seeking [36] and looking for excitement [35], [36]. Those individuals

seek power, status and recognition [37], therefore socializing with peers [38], [37], [39], [40] and building relationships for future necessity [41]. Introverts (which is the reverse of extraverts) prefer to be alone, and thus, they are less likely to be influenced by others' cheating behavior [42]. In contrast, extraverts are more impulsive and less self-controlled [43]. This tendency causes them to be more vulnerable to an unethical behavior as they are prone to imitate others [44]. This being said, it's important to note that the studies that addressed this trait's effect on academic dishonesty are scarce and their results are contradicting [34]. Some of them have found that extraversion was positively related to cheating behavior and academic dishonesty [45], [46], [47], while others did not find this impact on the level of cheating tendency among students [32].

Conscientiousness describes organized and responsible individuals, who think and plan before taking any action and follow society's rules and norms. Accordingly, the Conscientious student may be described as dependable, achievement-oriented, persistent, responsible and honest [39]. He operates as an effective regulator of his own actions, who is able to restrain and regulate behavior through "effortful control", thus, he can resist cheating [48] and hold more negative attitudes towards academic dishonesty [49]. He acts with high productivity and less deviance [51]. As opposed, student with lower conscientiousness is expected to be irresponsible, disorganized and impulsive. As a consequence, these characteristics might lead to poorer studying skills, which in turn might increase the tendency to cheat. Accordingly, research has found that this trait can foresee unethical behaviors [51], [52], [53].

Agreeableness involves cooperating with others and maintaining harmony. Individuals that are high on agreeableness have high ability to create good relationship [54] and peruse group norms [55]. In contrast, those that are low in this trait are expected to be lacking in these behaviors. Although research has not found a significant impact of agreeableness on academic dishonesty in general [32], the study of Williams et al. [34] found that agreeableness was negatively correlated with a particular unethical academic behavior of plagiarism. Thus, it can be expected that individuals who are high in agreeableness will follow the rules and norms and will less implement deviant behavior.

Neuroticism is a personality trait that describes an individual with a non-constructive emotionality [39], [33]. Thus, not surprisingly, in research literature it has been associated with organizational deviance [51] and has a negative impact on the tendency to engage in unethical academic behaviors and in cheating [34]. Emotional Stability (which is the reverse of neuroticism) reflects students' enhanced feeling of competence and a sense of security [39], which allows them to be more relaxed, unworried and less likely to become strained in stressful conditions, such as tests or deadlines. Thus, these students are considered to be less inclined toward cheating behaviors [49].

Finally, high Openness to Experience includes tendencies toward intellectualism, creativity, imagination, and broad-mindedness [39], [56], cognitive capability and high training aptitude [39]. Research findings show that this personality trait is related to academic success and to learning orientation, reflecting desire to understand concepts and master material [49]. Furthermore, learning orientation predicted lower inclination to cheat [50].

Empirical research confirmed the relationship between personality traits and individual tendency to cheat for Extraversion and for Neuroticism [30]. In addition, low Conscientiousness and low Agreeableness was found as predicting cheating behaviors as well [38]. More recently, Day et al. [49] examined the effects of Conscientiousness, Emotional Stability, and Openness to Experience on students' attitudes towards cheating, combined with two context variables, e.g., classroom culture and pedagogy. The findings showed that while

Conscientiousness was the sole personality measure that directly predicted negative attitudes towards cheating, Emotional Stability and Openness to Experience also lead to negative attitudes towards academic misconduct, however, only when combined with classroom context variables. Based on the above, we hypothesize that there will be differences in the level of academic dishonesty based on the various personality traits especially among e-learners.

Method

Participants

The sample consisted of 1,574 participants with 803 from two American academic institutes and 771 from four Israeli academic institutes. 65% of the participants were women and 35% were men. The age ranged from 17 to 59 (the mean was 26.4 years). 26% of the participants were freshmen, 32% - sophomores, 20% - juniors, 19% - seniors, and 3% were graduate students. 46% were Christians, 38% were Jews, and 16% were Muslims. 13% of the participants were excluded from the analysis because their surveys were incomplete or carelessly completed. Therefore, the final data set consisted of 1,365 participants.

Survey Instrument

A three part survey instrument was used in the current study. Part 1 included the TIPI scale developed by [59], which was consisted of 10 items assessing the participants' personality traits. The reliability of this questionnaire, measured by Cronbach’s alpha, was 0.72. Part 2 was consisted of the questions that examined academic integrity using the Academic Integrity Inventory [57]. These questions investigated the students’ likelihood to engage in various forms of academic misconduct. The instrument was validated by [57] and reliability of this questionnaire, measured by Cronbach’s alpha, was 0.75. Part 3 presented a series of socio-demographic questions.

Procedure

In order to encourage the participants to think in the frame of a specific type of course, we administered a printed version of the survey instrument in the traditional face-to-face courses and an on-line version of the survey instrument in the e-learning courses. The survey instruments were coded and grouped according to the location of the participants’ college or university (USA or Israel). The questionnaires were distributed at the end of the courses.

Results

Table 1 summarizes the results of Independent Sample T-test analyses, which indicate that there were statistically significant differences in students’ likelihood to engage in academic dishonesty based on the type of course in which they were enrolled. Specifically, it was found that students in face-to-face courses were more likely to engage in acts of academic dishonesty than their counterparts in e-learning courses.

Table 1

Differences in academic dishonesty by course type and country

Country	Course type	N	Mean	S.D.	T-Test	F
USA	E-learning	287	1.61	0.52	12.70^{***}	57.16^{***}
	Face-to-Face	470	2.16	0.66		
Israel	E-learning	293	1.78	0.60	5.33^{***}	
	Face-to-Face	315	2.52	0.65		

$^{***}P<0.001$, $^{**}P<0.01$, $^{*}P<0.05$

Based on multiple analysis of variance (MANOVA) analysis we found significant interaction between country and course type [$F_{(1, 1361)}$ =57.16, p<0.001].

Table 2

Correlation between personality and academic dishonesty by course type and country

Country	Course type	1	2	3	4	5
Israel N= 608	On-line	-0.038	-0.149*	-0.125*	-0.246**	-0.068
	Face to face	-0.090	0.131-*	-0.237**	-0.151**	-0.063
USA N= 757	On-line	-0.100	-0.090	0.057-	0.121-*	-0.038
	Face to face	-0.016	-0.040	0.031-	-0.114*	0.105*

***P<0.001, **P<0.01, *P<0.05

Note: 1=Openness to Experiences, 2=Emotional Stability, 3=Consciousness, 4=Agreeableness, 5=Extraversion

Table 2 shows a significant negative correlation between the personality trait of Agreeableness and academic dishonesty in both countries: Israel and USA. In addition, Israeli students identified with higher Conscientiousness and Emotional Stability demonstrated a significant negative correlation with academic dishonesty. General Linear Model revealed that there is a significant 2-way interaction effect among Israeli students between course type (on-line vs. face-to-face) and the personality trait of Conscientiousness [F=2.058, p<0.05] and between course type and the personality trait of Emotional Stability [F=2.047, p<0.05]. Interestingly, the personality trait of Extraversion among American students was found in a positive correlation with academic dishonesty, indicating that the tendency to be sociable is correlated with higher inclination to cheat.

Discussion and Conclusion

Our research found that there is less overall cheating in the virtual than in the traditional classroom settings. These findings are similar to [58] and [59], who explained this phenomenon by the notion that these students may have a higher motivation to learn or able to learn without being dependant on the typical structure of traditional classroom settings.

Our research also indicates that the personality traits of Emotional Stability and Conscientiousness are negatively related to academic dishonesty. These findings are similar to [49] and can be explained by to the notion that students with high Conscientiousness have the proper tendencies to be able to effectively regulate his actions and restrain inappropriate behaviors [48], [51], [52], [53]. Emotional Stability also leads students to be less inclined toward cheating behaviors [49] by enhancing feelings of competence and providing them with sense of security, which in result allow them to successfully cope with stressful situations and conditions [39]. In addition, a significant negative correlation between the personality trait of Agreeableness and academic dishonesty indicates that the more are the students cooperative with others, since the trait of Agreeableness is associated with the ability to create good relationship and to conform with group norms [54], [55], the less are they likely to be academically dishonest.

The results of this research showed that these effects were not observed among American students. This might be explained by cultural differences, as several studies that compared US students with students in Lebanon [60], China [61] and non-Western countries [62], showed that Americans tend to show less acceptance for cheating and to possess higher standards with regard to honesty.

The main practical implication and contribution of this research are to the process of students' profiling, since we found that students who use cheating practices are less emotionally stable, less conscious and less agreeable. Further research should focus on how to amplify cooperative tasks in online courses in order to reduce Academic Dishonesty. Classroom contextual effects, such as those presented in [49]'s study, may be worth investigating in further research as well, since they seem to contribute to the knowledge regarding the effects of personality traits on attitudes toward cheating [49].

References

[1] Becker, D. Connolly, J. Lentz, P. and Morrison, J. (2006) 'Using the business fraud triangle to protect academic dishonesty among business students', *Leadership Journal,* vol. 10, no. 1, pp.37-54, 2006.

[2] Bloodgood, J.M., Turnley, W.H., and Mudrack, P. (2008) 'The influence of ethics instruction, religiosity, and intelligence on cheating behavior', *Journal of Business Ethics,* vol. 82, pp.557-571.

[3] Sierra, J.J. and Hyman, M.R. (2008) 'Ethical antecedents of cheating intentions: Evidence of mediation', *Journal of Academic Ethics,* vol. 6, pp.51-66.

[4] Williams, M.S. and Hosek, W.R. (2003) 'Strategies for reducing academic dishonesty', *Journal of Legal Studies Education,* vol. 21, p.87.

[5] Anitsal, I.M., Anitsal, M. and Elmore, R. (2009) 'Academic dishonesty and intention to cheat: A model on active versus passive academic dishonesty as perceived by business students', *Academy of Educational Leadership Journal,* vol. 13, no. 2, pp.17-26.

[6] Park, C. (2003). 'In other (people's) words: Plagiarism by university students – literature and lessons', *Assessment and Evaluation in Higher Education,* vol. 28, pp.471-488.

[7] Pullin, R., Ortloff, V., Casey, S. and Payne, J.B. (2000) 'Analysis of academic misconduct using unobtrusive research: A study of discarded cheat sheets', *College Student Journal,* vol. 34, p. 616.

[8] Brown, B.S. and McInerney, M. (2008) 'Changes in academic dishonesty among business students in the United States, 1999-2006', *International Journal of Management,* vol. 25, no. 3, pp.621-632, 2008.

[9] Lambert, E.G. and Hogan, N.L. (2004) 'Academic Dishonesty among Criminal Justice Majors: A Research Note', *American Journal of Criminal Justice*, vol. 29, no. 1, pp. 1-20.

[10] Tibbetts, S. G. and D. L. Myers. (1999) 'Low Self-Control, Rational Choice, and Student Test Cheating', *American Journal of Criminal Justice*, vol. 23, no. 2, pp.179-201.

[11] Lanier, M. (2006) 'Academic Integrity and Distance Learning', *Journal of Criminal Justice Education,* vol. 17, no. 2, pp.244–261.

[12] Comas-Forgas, R. and Sureda-Negre, J. (2010) 'Academic plagiarism: Explanatory factors from students', *perspective. Journal of Academic Ethics*, vol. 8, no. 3, pp.217–232.

[13] Connors, M. (1996) 'Cybercheating: The Internet Could Become the Newest Battleground in Academic Fraud', *The Muse*. Retrieved from http://ursu.uregina.ca/_carillon/ Sept12.96/feature/feature1.html. Accessed November 1998.

[14] Wood, G. and Warnken P. (2004) 'Managing Technology: Academic Original Sin: Plagiarism, the Internet and Librarians', *Journal of Academic Librarianship*, vol. 30, no. 3, pp.1-7.

[15] Simkin, M.G. and McLeod, A. (2010) 'Why Do College Students Cheat?', *Journal of Business Ethics*, vol. 94, pp. 441– 453.

[16] Rosamond, B. (2002) 'Plagiarism, academic norms, and the governance of the profession', *Politics*, vol. 22, no. 3, pp.167-174.

[17] iNACOL. (2010) National standards for quality online teaching, Vienna, VA: Author. In

Jomon, A.P. and Cochran, J.D. (2013) 'Key interactions for online programs between faculty, students, technologies and educational institutions', *The Quarterly Review of Distance Education*, Vol. 14, no. 1, pp. 49–62.

[18] Davis, S. F., Kennedy, K., Nowak, S., Raghuramam, R. and Thomas, J. (2000) 'Academic Dishonesty and Distance Learning: Student and Faculty Views', *College Student Journal*, vol. 34, no. 2, pp. 309-15.

[19] Robinson-Zañartu, C., Peña, E.D., Cook-Morales, V., Peña, A.M., Afshani, R. and Nguyen, L. (2005) 'Academic crime and punishment: Faculty members’ perceptions of and responses to plagiarism', *School Psychology Quarterly*, vol. 20, no. 3, pp.318-337.

[20] Walker, J. (2010) 'Measuring plagiarism: Researching what students do, not what they say they do', *Studies in Higher Education,* vol. 35, no. 1, pp. 41–59.

[21] Kelley, K. and Bonner, K. (2005) 'Distance education and academic dishonesty: Faculty and administrator perception and responses', *Journal of Asynchronous Learning Network*, vol. 9, pp. 43-52.

[22] Burgoon, J., Stoner, M., Bonita, J. and Dunbar, N. (2003) 'Trust and Deception in Mediated Communication', 36th Hawaii International Conference on Systems Sciences, 44a.

[23] Rowe, N. (2004) 'Cheating in online student assessment: Beyond plagiarism', *Online Journal of Distance Learning*.

[24] Heberling, M. (2004) 'Maintaining academic integrity in online education', *Online Journal of Distance Learning Administration*, vol. 5.

[25] Kennedy, K., Nowak, S., Raghuraman, R., Thomas, J. and Davis, S.F. (2000) 'Academic dishonesty and distance learning: Student and faculty views', *College Student Journal*, vol. 34, pp. 309-314.

[26] Ford, R.C. and Richardson, W.D. (1994) 'Ethical decision making: A review of the empirical literature', *Journal of Business Ethics,* vol. 13, pp.205-221.

[27] Chapman, J.K., Davis, R., Toy, P. and Wright, L. (2004) 'Academic integrity in the business school environment: I’ll get by with a little help from my friends', *Journal of Marketing Education,* vol. 26, no. 3, pp.236-249.

[28] Allmon, D.E., Page, D. and Rpberts, R. (2000) 'Determinants of Perceptions of Cheating: Ethical Orientation, Personality and Demographics', *Journal of Business Ethics*, vol. 23, no. 4, pp.411-422.

[29] Etter, S., Cramer, J. and Finn, S. (2006) 'Origins of Academic Dishonesty: Ethical Orientations and Personality Factors Associated with Attitudes about Cheating with Information Technology', *Journal of Research on Technology in Education*, vol. 39, no. 2, pp. 133-155.

[30] Jackson, C. J., Levine, S. Z., Furnham, A. and Burr, N. (2002) 'Predictors of Cheating Behavior at a University: A Lesson From the Psychology of Work', *Journal of Applied Social Psychology*, vol. 32, pp. 1031-1046.

[31] Sackett, P.R. and Wanek, J.E (1996) 'New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection', *Personnel Psychology*, vol. 42, pp. 787- 829.

[32] Christine, Z.J. and James, C.A. (2008) 'Personality Traits and Academic Attributes as Determinants of Academic Dishonesty in Accounting and Non-Accounting College Majors', In the Proceeding of 15th Annual Meeting of American Society of Business and Behavioral Sciences (ASBBS) held on 23-24 February, 2008 at Las Vegas, USA, 15, 604-616.

[33] John, O. P. and Srivsatava, S. (1999) 'The Big Five Trait Taxonomy: History, Measurement and Theoretical Perspectives', In Pervin L and OP Johns (Eds.), Handbook of personality: Theory and Research, pp. 102-138, New York: Guilford.

[34] Williams, K.M., Nathanson, C. and Paulhus, D.L. (2010) 'Identifying and Profiling Scholastic Cheaters: Their Personality, Cognitive Ability & Motivation', *Journal of Experimental Psychology*, vol. 16, pp. 293-307.

[35] Digman, J. M. (1990) 'Personality Structure: Emergence of the Five-Factor Model', *Annual Review of Psychology*, vol. 41, pp. 417–440.

[36] Goldberg, L.R. (1993) 'The Structure of Phenotypic Personality Traits', *American Psychologist*, vol. 48, pp. 26–34.

[37] Costa, P.T. Jr. and McCrae, R.R. (1988) 'Personality in Adulthood: A Six Year Longitudinal Study of Self-Reports and Spouse Ratings on the NEO Personality Inventory', *Journal of Personality and Social Psychology*, vol. 54, pp. 853–863.

[38] Costa, P.T. Jr. and McCrae, R.R. (1992) Revised NEO Personality Inventory and NEO Five-Factor Inventory Professional Manual (Psychological Assessment Resources, Odessa, FL).

[39] Barrick M.R. and Mount, M.K. (1991) 'The Big Five Personality Dimensions and Job Performance: A Meta-Analysis', *Personnel Psychology*, vol. 44, pp. 1-26.

[40] Furnham, A., Petrides, K.V., Tsaousis, I. Pappas, K. and Garrod, D. (2005) 'A Cross-Cultural Investigation into the Relationships between Personality Traits and Work Values', *The Journal of Psychology*, vol. 139, pp. 5-32.

[41] Watson, D. and Hubbard, B. (1996) 'Adaptational Style and Dispositional Structure: Coping in the Context of the 5-Factor Model', *Journal of Personality*, vol. 64, pp. 737–774.

[42] O'Fallon, M.J. and Butterfield, K.D. (2011) 'Moral Differentiation: Exploring Boundaries of the ''Monkey See, Monkey Do'' Perspective', *Journal of Business Ethics*, vol. 102, pp. 379–399.

[43] Eysenck, H.J. and Eysenck M.W. (ed). (1985) Personality and individual differences: A natural science approach. New York: Plenum Press.

[44] Phillips, A.S. and Bedeian, A.G. (1994) 'Leader-Follower Exchange Quality: The Role of Personal and Interpersonal Attributes', *Academy of Management Journal*, vol. 37, no. 4, pp. 990–1001.

[45] Aslam, M.S. and Nazir, M.S. (2011) 'The impact of personality traits on academic dishonesty among Pakistan students', *The Journal of Commerce*, Vol. 3, No. 2.

[46] Cizek, G. J. (1999) 'Cheating on Tests: How to Do It, Detect It, and Prevent It', Erlbaum, Mahwah, NJ.

[47] Singh, U. and Akhtar, S.N. (1972) 'Personality variables and cheating in examinations', *India: Tata Institute of Social Sciences*, vol. 32, pp. 423-428.

[48] Jensen-Campbell, L.A. and Graziano, W.G. (2005) 'The two faces of temptation: Differing motives for self-control', *Merrill-Palmer Quarterly*, vol. 51, no. 3, pp. 287–314.

[49] Day, N.E., Hudson, D., Dobies, P.R. and Waris, R. (2011) 'Student or situation? Personality and classroom context as predictors of attitudes about business school cheating', *Social Psychology of Education*, vol. 14, no. 2, pp. 261-282.

[50] Davis, S., Pierce, M. C., Yandell, L. R., Arnow, P. S., and Loree, A. (1995). 'Cheating in college and the type a personality: A reevaluation', *College Student Journal*, vol. 29, no. 4, pp. 493–497.

[51] Colbert A.E., Mount, M.K., Harter, J.K., Witt, L.A. and Barrick, M.R. (2004) 'Interactive Effects of Personality and Perceptions of the Work Situation on Workplace Deviance', *Journal of Applied Psychology*, vol. 89, pp. 599-609.

[52] Litzky, B.E., Eddleston, K.A. and Kidder, D.L. (2006) 'The Good, the Bad, and the Misguided: How Managers Inadvertently Encourage Deviant Behaviors', *Academy of Management Perspectives*, vol. 20, pp. 91-103.

[53] Robinson, S.L. and Greenburg, J. (1998) 'Employees Behaving Badly: Dimensions, Determinants, and Dilemmas in the Study of Workplace Deviance', In C.L. Cooper and D.M. Rousseau (Eds.), Trends in organizational behavior, pp. 1-30. Chichester: John Wiley & Sons.

[54] Organ, D.W. and Lingl, A. (1995) 'Personality, Satisfaction, and Organizational Citizenship Behavior', *The Journal of Social Psychology*, vol. 135, pp. 339-350.

[55] Trevino, L.K. and Nelson, K.A. (2004) Managing Business Ethics: Straight Talk about How to do it Right (3rd Ed.). Hoboken, NJ: John Wiley & Sons, Inc.

[56] Gosling, S.D., Rentfrow, P.J. and Swann, W.B., Jr. (2003) 'A very brief measure of the Big Five personality domains', *Journal of Research in Personality*, vol. 37, pp. 504-528.

[57] Kisamore, J.L., Stone, T.H. and Jawahar, I.M. (2007) 'Academic integrity: The relationship between individual and situational factors on misconduct contemplations', *Journal of Business Ethics*, vol. 75, pp. 381–394.

[58] Stuber-McEwen, D., Wiseley, P. and Hoggatt, S. (2009) 'Point, Click, and Cheat: Frequency and Type of Academic Dishonesty in the Virtual Classroom', 2009.

[59] Peled, Y., Barczyk, C., Eshet, Y., and Grinautski, K. (2012) 'Learning Motivation and Student Academic Dishonesty – A Comparison between Face-To-Face and Online Courses', In P. Resta (Ed.), Proceedings of Society for Information Technology & Teacher Education International Conference 2012 (pp. 752-759). Chesapeake, VA: AACE.

[60] McCabe, D.L., Feghali, T. and Abdallah, H. (2008) 'Academic Dishonesty in the Middle East: Individual and Contextual Factors', *Research in Higher Education*, vol. 49, no. 5, pp. 451-467.

[61] Rawwas, M., Al-Khatib, J. and Vitell, S. (2004) 'Academic Dishonesty: A Cross-Cultural Comparison of U.S. and Chinese Marketing Students', *Journal of Marketing Education*, vol. 26, pp. 89-100.

[62] Grimes, P.W. (2004) 'Dishonesty in Academics and Business: A Cross-Cultural Evaluation of Student Attitudes', *Journal of Business Ethics*, vol. 49, pp. 273-290.

DAVID PUBLISHING

Statistical Science and Application
Volume 2, Number 3, March 2014

David Publishing Company
240 Nagle Avenue #15C, New York, NY 10034, USA
Tel: 1-323-984-7526, 323-410-1082; Fax: 1-323-984-7374, 323-908-0457
http://www.davidpublishing.com
statistical@davidpublishing.org, jssa.statistical@hotmail.com

ISSN 2328-224X
03